PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the Application of:

REBECCA E. CAHOON ET AL.

APPLICATION NO.: 10/659,869 GROUP ART UNIT: 1638

FILED: SEPTEMBER 11, 2003 EXAMINER: MEDINA AHMED IBRAHIM

DOCKET NO.: BB1294USCNT CONFIRMATION NO.: 6089

FOR: PLANT MYB TRANSCRIPTION FACTOR HOMOLOGS

Declaration Pursuant to 37 CFR §1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

I, Joan T. Odell, am a citizen of the United States of America, residing at 127 Monitor Place, Unionville, Pennsylvania, United States of America, and I declare as follows:

1. I am one of the co-inventors named in above-identified application.

2. I received a B.A. degree in Biology from the University of California at San Diego in 1975. I received a Ph.D. degree in Biology from the University of California at San Diego in 1981. I was a Postdoctoral Fellow at the Rockefeller University from 1981 to 1985.

3. I have been employed by E. I. du Pont de Nemours and Company from 1985 to the present. From 1985 to 2001, I was a Principle Investigator, conducting and directing research in the areas of plant gene expression and genetic engineering. From 2001 to 2003, I served as a Six Sigma specialist. From 2003 to the present, I have served as a patent liaison. I became a registered patent agent

with the United States Patent and Trademark Office in 2005 (Registration Number 56,870).

4. I have reviewed the Office Action dated October 19, 2005. I am aware that this declaration is being submitted to illustrate work which I, or those working under my guidance, have done to demonstrate that the polypeptide set forth in SEQ ID NO:36 (encoded by SEQ ID NO:35) functions as a Myb-related transcription factor in a transgenic cell or plant.

5. An expression cassette was constructed for expression of the soybean Myb-related transcription factor, "soyMyb2 " (SEQ ID NO:36; encoded by SEQ ID NO:35). The soyMyb2 expression cassette is comprised the following elements:

(a) CaMV 35S promoter;

(b) Modified protein-coding region of SEQ ID NO:35; and

(c) Nopaline synthase (*nos*) 3' end.

The soyMyb2 expression cassette was constructed in the following manner: The soyMyb2 sequence was isolated by PCR amplification of the clone, sfl1.pk0105.e6, using the following two primers:

(a) Forward Primer-1: CACAAGTTCATGAATAAAAAACAAC; and

(b) Reverse Primer-2: CAAACCCAATAATATGTTTTAA.

The Forward Primer-1 introduced a BspHI site, 5'-TCATGA, into the soyMyb2 sequence, which overlapped the start methionine codon (ATG). This primer produced a point mutation in the soyMyb2 protein-coding region (PCR product). The second amino acid of soyMyb2 was changed from aspartic acid to asparagine (GAT -> AAT). The deduced nucleotide sequence of the soyMyb2 PCR product is presented in Appendix B, which accompanies this paper. This nucleotide sequence is comprised of the following segments: nucleotides 1-25, which correspond to the nucleotide sequence of Forward Primer-1 (primer nucleotides 15-25 correspond to nucleotides 34-44 of SEQ ID NO:35); and nucleotides 26-709, which correspond to nucleotides 45-728 of SEQ ID NO:35, which includes nucleotides 688-709, which correspond to the reverse complement of the nucleotide sequence of Reverse Primer-2.

The soyMyb2 PCR product was cloned into the vector, pCR™2.1-TOPO, using TOPO™ TA. The resulting plasmid was called myb2-topo2.1.

A nucleic acid fragment encoding soyMyb2 was then isolated from the plasmid, myb2-topo2.1, using restriction enzymes BspHI and KpnI. The KpnI cleavage site is within the pCR™2.1-TOPO vector sequence. The BspHI-KpnI

fragment containing the soyMyb2 coding region was then cloned into plasmid DNA pMH40, which had been restriction digested with NcoI and KpnI (the BspHI and NcoI 5'-overhangs are identical). The modified soyMyb2 PCR product replaced the GUS coding region that was present in the pMH40 expression vector. The resulting plasmid was called myb305-2-pMH40.

In plasmid myb305-2-pMH40, a CaMV 35S promoter was used to drive gene expression of the soyMyb2 PCR product. The CaMV 35S promoter is a 1337-bp fragment, as illustrated in Appendix C, accompanied herewith. The nos 3'-end used to regulate transcription termination is a 761-bp fragment, as illustrated in the accompanying Appendix D.

The soyMyb2 expression cassette from plasmid myb305-2-pMH40 was cloned into a binary vector, pZBLN1N. The binary vector pZBLN1N contains right and left T-DNA borders, as well as plant and bacterial expression cassettes, each containing a neomycin phosphotransferase II ("nptII") gene for selection using kanamycin. The resulting vector, myb305-2-pZBL1N, was transformed into *Arabidopsis*.

Arabidopsis that were successfully transformed with the soyMyb2 expression cassette produced purple seedlings. The cotyledons, hypocotyls, and older parts of the roots also displayed purple color. The adult plants looked normal but exhibited slower growth and produced fewer and poorer seeds.

The red and purple pigmentation in plant tissues is due to secondary metabolites called anthocyanins, which are produced through the flavonoid pathway, a branch of the general phenylpropanoid pathway. As an example, this pathway is discussed in Uimari et al., 1997 *Plant J* 12:1273-1284. A copy of Uimari et al. is provided herewith.

These results indicated that expression of the soyMyb2 chimeric gene activated the anthocyanin pathway in transgenic *Arabidopsis*.

6. An expression cassette was constructed for expression of a chimeric Myb-related transcription factor, "soyMyb2-PvALF ". PvALF (GI No. 1046278) is an ABI-like transcription factor from *Phaseolus vulgaris*. The PvALF transcription factor has a transcription activation domain at the amino-terminus (Bobb et al., 1995 *Plant J* 8:101-113). The soyMyb2-PvALF expression cassette comprised the following elements:

(a) CaMV 35S promoter;
(b) Nucleotides 29 - 442 of SEQ ID NO:35, which encode the DNA-binding domain of soyMyb2 (amino acids 1 - 138 of SEQ ID NO:36);

(c) 241 amino acid transcriptional activation domain of PvALF; and

(d) Nopaline synthase (*nos*) 3'-end.

The soyMyb2-PvALF expression cassette was constructed in the following manner:

The DNA-binding domain of the soyMyb2 gene was isolated by PCR amplification of the clone sfl1.pk0105.e6, using the following primers:

(a) Forward Primer-3:
TGTCACCATGGATAAAAAACAACAGTGTAAGACGTC;

(b) Reverse Primer-4:
TTTGGACCCCGGGAATTCGTGATCATTTATCTCAGAATTATTACTAC TC

The Forward Primer-3 provided a NcoI recognition site (5'-CCATGG), overlapping the start methionine codon. The Reverse Primer-4 provided SmaI (5'-CCCGGG) and EcoRI (5'-GAATTC) recognition sites. The deduced nucleotide sequence of the resulting PCR product, DNA-binding domain of SoyMyb2, is presented in Appendix E, a copy of which is attached hereto. This nucleotide sequence is comprised of the following segments: nucleotides 1-36, which correspond to the nucleotide sequence of Forward Primer-3; nucleotides 37-390, which correspond to nucleotides 58-411 of SEQ ID NO:35; and nucleotides 391-439, which correspond to the reverse complement of the sequence of Reverse Primer-4.

The PCR product was cloned into the Promega pGEM™-T Easy vector (AT-tailed). The resulting plasmid DNA was called pGMBD-5.

The plasmid 108G4Alf contains the phaseolin promoter, Gal4-PvALF fusion protein, and phaseolin 3'-end. The use of the transcriptional activation domain of PvALF in the construction of chimeric transcription factors has been previously described (U.S. Patent No. 5,968,793; Example 2). A nucleic acid fragment from pGMBD-5 that contained the soyMyb2 DNA-binding domain was obtained by restriction digesting with NcoI and EcoRI. This fragment was cloned into plasmid 108G4Alf that also had been restriction digested with NcoI and EcoRI. This resulted in replacement of both the Gal4 DNA-binding domain and the 5' region of the PvALF activation domain (an EcoRI-EcoRI fragment) of plasmid 108G4Alf with the soyMyb2 DNA-binding domain. The resulting plasmid was named p108MBD. It contained the intact soyMyb2 DNA-binding domain and the 3' region of the PvALF transcription activation domain.

Plasmid p108MBD was restriction digested with EcoRI. The 5' region of the PvALF activation domain was inserted as an EcoRI-EcoRI fragment from the

plasmid 35S-G4Alf. The resulting plasmid DNA, with an intact soyMyb2 DNA-binding domain and an intact PvALF activation domain, was called 108MybA.

The deduced nucleotide sequence of the chimeric soyMyb2-PvALF protein-coding region is presented in Appendix F, a copy of which accompanies this paper. This nucleotide sequence is comprised of the following segments: nucleotides 1-420, which correspond to nucleotides 8-427 of the soyMyb2 DNA-binding domain PCR product (presented in Appendix E); nucleotides 421-426, which correspond to the SmaI linker preceding PvALF in plasmid p108G4Alf; nucleotides 427-1129, which correspond to nucleotides 41-743 of GI No. 1046277 (PvALF); nucleotides 1130-1155, which correspond to the reverse complement of nucleotides 4-29 of the PvALF PCR primer, Alf6-SalI (described in Bobb *et al.*, 1995 *Plant J 8*:101-113); nucleotides 1156-1167, which correspond to nucleotides 2086-2097 of the linker region of plasmid p108G4Alf; and nucleotides 1168-1176, which correspond to nucleotides 3287-3295 of plasmid pML63, the sequence that immediately precedes the nos 3'-end.

Plasmid pML63 contains an expression cassette containing the following elements: a CaMV 35S promoter; a GUS coding sequence; and a nos 3'-end. The GUS coding region of pML63 was replaced with the soyMyb2-PvALF coding region in the following manner. The soyMyb2-PvALF coding region was isolated from plasmid 108MybA as two restriction fragments, an NcoI-BglII fragment (for the 5' region) and a BglII-SmaI fragment (for the 3' region). These two fragments were cloned into plasmid pML63 that had been digested with NcoI and SmaI. The resulting plasmid, containing a soyMyb2-PvALF expression cassette, was called p35MybA.

The CaMV 35S promoter present in p35MybA is a 1404 nucleotide fragment, which is presented in Appendix G, accompanied herewith.

The nos 3'-end region present in p35MybA is a 279 nucleotide fragment, which is presented in Appendix H, accompanied herewith.

An XbaI fragment from p35MybA was cloned into the binary vector pZBLN1N. The resulting plasmid, pB35MybA, was used to transform *Arabidopsis*.

Arabidopsis that were successfully transformed with the soyMyb2-PvALF expression cassette produced purple seedlings. This purple color was slightly more intense than that observed for the transgenic soyMyb2 seedlings described above. The cotyledons, hypocotyls, and older parts of the roots were all purple. The adult plants looked normal but exhibited slower growth and produced fewer and poorer seeds.

These results indicated that expression of the novel soyMyb2-PvAlf chimeric gene activated the anthocyanin pathway in transgenic *Arabidopsis*.

In summary, the results of the transgenic *Arabidopsis* studies with soyMyb2 and the chimeric polypeptide soyMyb2-PvALF is believed to show that the polypeptide set forth in SEQ ID NO:36 exhibits Myb-related transcription factor activity. The activation of the anthocyanin pathway in the transgenic *Arabidopsis* plants was evidenced by the purple color of the *Arabidopsis* seedlings produced by the transformed *Arabidopsis* plant.

I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Joan T. Odell

Joan T. Odell

April 18, 2006

Date

The Plant Journal (1995) 8(3), 331–343

PvAlf, an embryo-specific acidic transcriptional activator enhances gene expression from phaseolin and phytohemagglutinin promoters

Andrew J. Bobb, Hans Georg Eiben and Mauricio M. Bustos*
Department of Biological Sciences, UMBC, 5401 Wilkens Ave., Baltimore, MD 21228-5398, USA

Summary

Mutations in *Vp1* and *ABI3* genes of maize and *Arabidopsis* lead to drastic reductions in the synthesis of a subset of maturation-specific products including seed storage proteins. Gene *Phaseolus vulgaris* ABI3-like factor (*PvAlf*), whose protein product is similar to the ABI3 and Vp1 proteins, has been cloned. Here, it is shown that PvAlf positively regulates phaseolin and phytohemagglutinin (PHA-L) promoters in particle bombardment assays. PvAlf mRNA expression is embryo-specific and temporally complex. PvAlf mRNA abundance is highest during two periods (9–14 and 22–35 days after flowering) that precede the onsets of seed maturation and seed abscission, respectively. Protein fusions with the DNA-binding domain of the yeast transcriptional activator GAL4 demonstrated that the N-terminal 243 amino acids of PvAlf function as a strong transcriptional activation domain in yeast (*Saccharomyces cerevisiae*) and plant cells. This domain consists of a central cluster rich in serine, threonine and proline (STP cluster) flanked by two negatively charged regions containing bulky hydrophobic residues similar to acidic activation domains of Vp1, the herpes simplex virus virion protein VP16 and transcription factors GCN4 and HAP4 from yeast. Together with the Vp1 proteins of maize and rice and ABI3, PvAlf constitutes a class (Vp1/ABI3-like factors or VAlfs) of regulatory factors that are pivotal for the promotion of seed maturation and dormancy in angiosperms.

Introduction

When cotyledon-stage embryos of many plant species are cultured in water or in simple nutrient solutions they undergo germinative changes normally seen only during rehydration (imbibition) of mature, dry seeds (reviewed in Crouch, 1987; Galau *et al.*, 1991). Initial changes can occur in the presence of inhibitors of RNA synthesis (Long *et al.*, 1981) suggesting that plant embryos are already capable of germinating by the end of the cotyledon stage. However, precocious germination *in planta* (vivipary) is rare; instead, embryogeny proceeds into a maturation phase characterized by abundant expression of a limited set of specific genes (*MAT* genes) encoding storage proteins, lectins, oil body proteins, enzymes involved in lipid and starch metabolism, desiccation protectants and defense enzymes (glucanases, chitinases, amylases, etc.). In angiosperms (monocots and dicots), normal seed maturation and subsequent dormancy are disrupted by *aba* and *viviparous* mutations that affect the biosynthesis of the phytohormone abscisic acid (Koorneef *et al.*, 1982; Neill *et al.*, 1986; Robertson, 1955). Mutations that reduce the sensitivity of seed tissues to abscisic acid, such as *abi3* in *Arabidopsis thaliana* and *vp1* in maize, exhibit similar, albeit complex, developmental phenotypes that include premature germination (Robertson, 1955), insensitivity to abscisic acid (Koorneef *et al.*, 1984) and reduced storage protein accumulation (Kriz *et al.*, 1990; Nambara *et al.*, 1992) among other traits. However, *vp1* alleles show a deficiency in pigmentation (Hattori *et al.*, 1992; Neill *et al.*, 1986; Robertson, 1955) that is not observed in *abi3* mutants and, therefore, it is not clear whether they represent the same gene. Most relevant to this work are the dramatic effects that mutations in *ABI3* and *Vp1* genes have on *MAT* gene expression (Koorneef *et al.*, 1989; Nambara *et al.*, 1992, 1994; Paiva and Kriz, 1994; Pang *et al.*, 1988; Parcy *et al.*, 1994; Pla *et al.*, 1991); the strongest *abi3* mutant alleles (*abi3-4* and *abi3-6*) cause a near complete loss of 2S and 12S storage protein expression in *Arabidopsis* seeds (Nambara *et al.*, 1994; Parcy *et al.*, 1994), and a similar phenotype has been observed with respect to expression of a maize globulin gene (*Glb1*) in *vp1* null seeds (Kriz *et al.*, 1989). The *ABI3* gene was isolated by map-based positional cloning (Giraudat *et al.*, 1992) and shown to encode a protein similar to the product of *Vp1* (McCarty *et al.*, 1991). Vp1 and the equivalent gene from rice, OsVp1, can activate gene expression from Em and C1 promoters in maize endosperm protoplast assays (Hattori *et al.*, 1992, 1994; McCarty *et al.*, 1991). The transcriptional activation domain of Vp1 was localized within the first 121 amino acids at the N-terminus (McCarty *et al.*, 1991). Although similar information has not been reported for ABI3, Parcy *et al.* (1994) showed that ectopic expression of ABI3 and exposure to high levels of ABA in transgenic *Arabidopsis* leaves leads to activation of genes that are normally expressed only in siliques, also consistent with a role for ABI3 as a transcriptional activator.

Received 16 January 1995; revised 19 May 1995; accepted 5 June 1995.
For correspondence (fax + 1 410 455 3875).

In bean, phaseolin (a 7S seed storage protein) and phytohemagglutinin (PHA, a lectin) are the most abundantly expressed seed proteins, together accounting for 70–80% of the total protein content of a mature embryo. 7S and 11S storage proteins are also major constituents of soybean, pea and *Vicia faba* seeds (Casey *et al.*, 1986). Lectins, too, represent important components of many dicot seeds and some lectin-related proteins have been found to possess insect deterring properties (Osborn *et al.*, 1988). Phaseolin and lectin genes are known to be seed-specific, appear to be coordinately regulated during maturation (Murray and Kennard, 1984; Staswick and Chrispeels, 1984), and provide useful molecular markers to investigate the control of seed maturation in bean. Here we report cloning gene *Phaseolus vulgaris* ABI3-like factor (*PvAlf*), whose protein product is related to ABI3 and Vp1. PvAlf activates both phaseolin and PHA promoters in cotyledon cells. PvAlf mRNA was found to be embryo-specific and expressed during two different periods of embryogeny: the first one precedes the induction of phaseolin and PHA expression and the onset of maturation, and the second coincides with predesiccation and desiccation stages. Ectopic expression of PvAlf in leaves is sufficient to activate transient expression from phaseolin and PHA promoters suggesting that other seed-specific factors are not required for PvAlf-mediated activation. When bound to a nearby promoter site, PvAlf activates transcription via a complex N-terminal, acidic domain. These results demonstrate a clear role for PvAlf as a positive transcriptional regulator of maturation-specific genes in bean and for related VAlfs in legumes.

Results

Cloning of Phaseolus vulgaris *ABI3-like factor from French bean*

Evidence from physiological, genetic and molecular studies points to the involvement of the abscisic acid insensitive-3 (*ABI3*) gene of *A. thaliana* in the promotion of seed maturation and its specific program of gene expression (Finkelstein and Sommerville, 1990; Koorneef *et al.*, 1984). A method based on the rapid amplification of cDNA ends (3'-RACE) technique (Frohman *et al.*, 1988) was used to clone an ABI3-like mRNA from developing bean embryos. Briefly, degenerate oligonucleotide primers were synthesized corresponding to the peptides MEDIGT and VWNMRY conserved in ABI3, Vp1 and OsVp1. Total cellular poly(A)$^+$ RNA from mid-maturation bean embryos was reverse-transcribed using an oligo-dT primer and Superscript™ (BRL-GIBCO) reverse transcriptase and the resulting cDNA mixture was used as template for two nested polymerase chain amplification reactions with *Taq* DNA polymerase and the degenerate primers (Experimental procedures). The products from the second amplification were directly cloned into a plasmid vector (pCR2000, Invitrogen) and those encoding an ABI3-like open reading frame (ORF) were identified by di-deoxy sequencing. The middle portion of the ABI3-like ORF was amplified using a degenerate oligonucleotide that corresponded to the conserved pentapeptide LPDFP found near the N-termini of Vp1 and ABI3, and two non-degenerate, gene-specific primers selected from the sequence of the 3'-RACE clone. After two nested PCR reactions a product with the expected size, 1.8 kbp, was obtained, cloned and sequenced. Finally, the 5'-end of the ABI3-like ORF was obtained by 5'-RACE (Frohman *et al.*, 1988), this time using gene-specific primers deduced from the intermediate, 1.8 kbp clone. After identifying sequences that appeared to include a translation start codon, new 5' and 3' gene-specific primers were synthesized and used to re-amplify the complete ORF from fresh poly(A)$^+$ RNA yielding clone pPvAlf. The longest open reading frame in pPvAlf encodes a 752 amino acid protein termed *Phaseolus vulgaris* ABI3-like factor (PvAlf). One of the partial cDNA clones isolated had a 9 bp duplication (arrowhead, Figure 1) that extends the PvAlf ORF by 3 amino acids and is likely to represent a different PvAlf allele. Only three single base pair differences were recorded out of more than 5.4 kbp of redundant PvAlf sequences. Even if all three were artifacts of PCR, an unlikely coincidence since they are all silent mutations, this result still indicates a high degree of accuracy (greater than 99.8% identity) in the reported PvAlf sequence, similar to those in public nucleotide databases. In Figure 1, the sequence of PvAlf (from clone pPvAlf) is aligned to ABI3 (AtABI3), maize Vp1 (ZmVp1) and rice Vp1 (OsVp1). The alignment was done using the programs PILEUP and GAP (University of Wisconsin Genetics Computer Group). Overall, pair-wise identities for the group were low: PvAlf–ABI3 = 48%, PvAlf–Vp1 = 38% and PvAlf–OsVp1 = 41%. As noted before (Giraudat *et al.*, 1992; Hattori *et al.*, 1994), the identity is concentrated on four domains corresponding to positions 80–126 (domain I), 281–338 (domain II), 533–547 (domain III) and IV, 655–772 (domain IV) on Figure 1. We propose naming this class of proteins Vp1/ABI3-like factors (VAlf), after Vp1 and ABI3, the first two members to be cloned. VAlf domain IV is the most conserved (84% overall sequence identity) suggesting that it performs an important function in monocot and dicot species. In Vp1, domain I is included within an N-terminal transcriptional activation domain. The homology of this region to corresponding portions of PvAlf and ABI3 is restricted to the runs of multiple serine residues.

PvAlf is encoded by a single-copy gene

In *Arabidopsis, ABI3* is a single-copy gene expressed only in seeds (Giraudat *et al.*, 1992; Parcy *et al.*, 1994). We

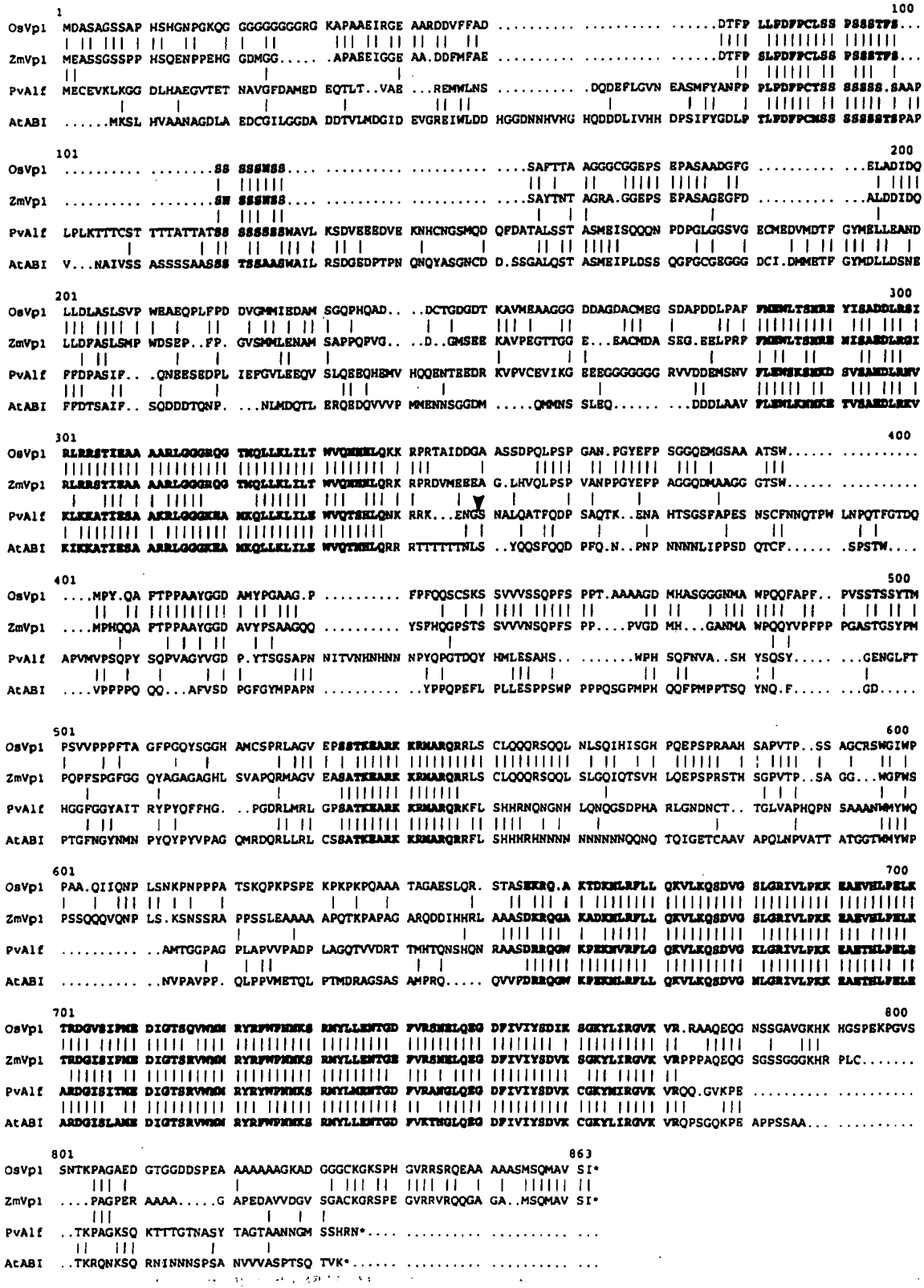

Figure 1. Gene *PvAlf* encodes a protein similar to the Vp1 and ABI3 factors of maize and *Arabidopsis*.
The deduced amino acid sequence of PvAlf was aligned to the sequences of ABI3 (Giraudat *et al.*, 1992), maize Vp1 (ZmVp1, McCarty *et al.*, 1991) and rice Vp1 (OsVp1, Hattori *et al.*, 1994) using the computer program PILEUP (University of Wisconsin Genetics Computer Group). Identical amino acids are indicated with vertical bars. The most conserved areas (domains I–IV) are highlighted in boldface type. The location of a 9 bp insertion coding for the tripeptide 'SNN' found in a different PvAlf clone is indicated by an arrowhead.

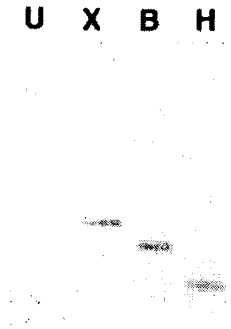

Figure 2. PvAlf is encoded by a single-copy gene.
Total genomic DNA isolated from etiolated bean seedlings (Experimental procedures) was size-fractionated by electrophoresis on a 0.75% agarose/TAE gel either uncut (U) or after restriction digestion with *Xba*I (X), *Bam*HI (B) or *Hin*dIII (H) After transfer to a nylon membrane, PvAlf sequences were detected by hybridization to a PvAlf ^{32}P-labeled probe and autoradiography.

investigated the number of PvAlf genes in bean using blot hybridization to genomic DNA digested with the restriction enzymes *Xba*I, *Bam*HI and *Hin*dIII. A sample of uncut DNA was included as a control. DNA blots were hybridized to a radiolabeled PvAlf probe under high-stringency conditions. The autoradiograph on Figure 2 shows the presence of a single *PvAlf* gene copy in bean. An additional cross-hybridizing band was detected at lower stringency, although it is unclear whether it represents a second gene bearing a low degree of homology to *PvAlf*.

PvAlf expression is seed-specific and developmentally regulated during embryogeny

The organ and temporal distribution of *PvAlf* mRNA expression were analyzed by RNA blot hybridization. Total cellular RNA was isolated from leaves, roots, seed pods, callus and eight stages of 'cotyledons' (cotyledons plus embryonic axis) ranging from early cotyledon to pre-desiccation collected between 9 and 35 days after the opening of the corollas (days after flowering or DAF). Self-pollination occurs in bean before the corollas open. The same filter was hybridized sequentially to PvAlf, phaseolin, phytohemagglutinin (PHA-L) and 18S rRNA (rRNA) probes. The relevant portions of each autoradiograph are shown in Figure 3. PvAlf mRNA expression was detected only in cotyledons. Even after prolonged exposure no signals could be seen in the lanes for leaf (L), root (R), seed pod (P) and callus suspension culture (C), all of which were loaded with the same amount of RNA (see rRNA panel). PvAlf mRNA was also absent from senescing leaves and flowers (data not shown). The developmental series of cotyledon stages revealed that the steady-state level of PvAlf mRNA was modulated during embryogeny. After the 9 DAF stage, PvAlf mRNAs segregated into two different

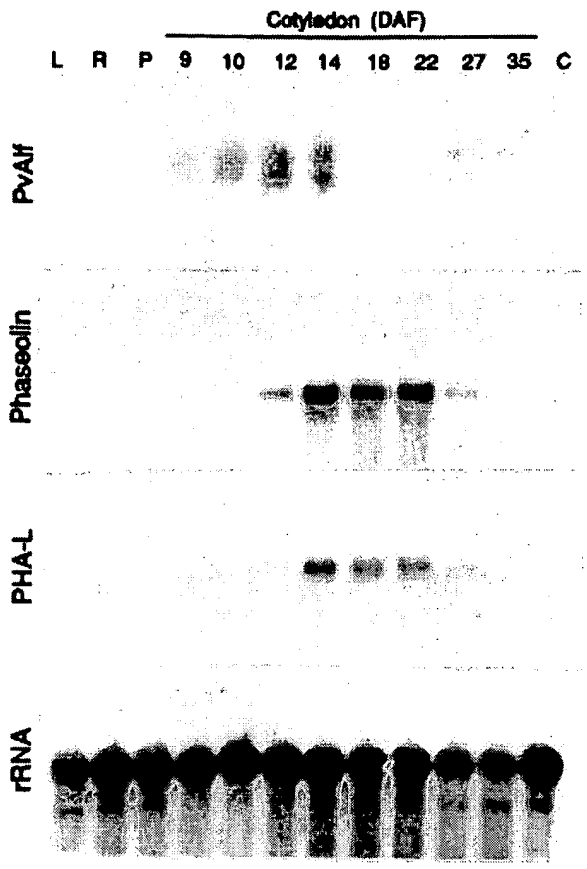

Figure 3. PvAlf expression is embryo-specific and developmentally regulated during embryogeny.
Total cellular RNA was isolated from leaves (L), roots (R), seed pods (P), callus suspension culture (C) and 'cotyledons' (cotyledons plus embryonic axis) at 9, 10, 12, 14, 18, 22, 27 and 35 days after flowering (DAF). The same amount of each RNA (10 μg per lane) was separated on a formaldehyde denaturing agarose gel, transferred to a nylon membrane and sequentially hybridized to ^{32}P-labeled probes specific for PvAlf, phaseolin, PHA-L and 18S rRNA. Exposure times varied from 2–4 h for phaseolin, PHA and 18S rRNA to 4 days for PvAlf.

size classes which increased in their abundance until approximately 14 DAF. This was followed by an interval of decreased expression lasting at most 6 days. A second period of increased expression was observed between 22 and 36 DAF. By contrast, the maturation markers phaseolin and PHA-L were induced at 10 DAF and their mRNAs were highest during the 14–22 DAF period. This is in agreement with expression profiles reported previously for both genes (Murray and Kennard, 1984; Staswick and Chrispeels, 1984). From this analysis we conclude that PvAlf mRNA expression is regulated in a complex manner during bean embryogeny. The first wave of PvAlf mRNA induction precedes the onset of maturation-specific gene expression by at least 2–3 days; a second wave of PvAlf mRNA expression occurs later in embryogeny, after the abundances of maturation-specific mRNAs begin to decrease.

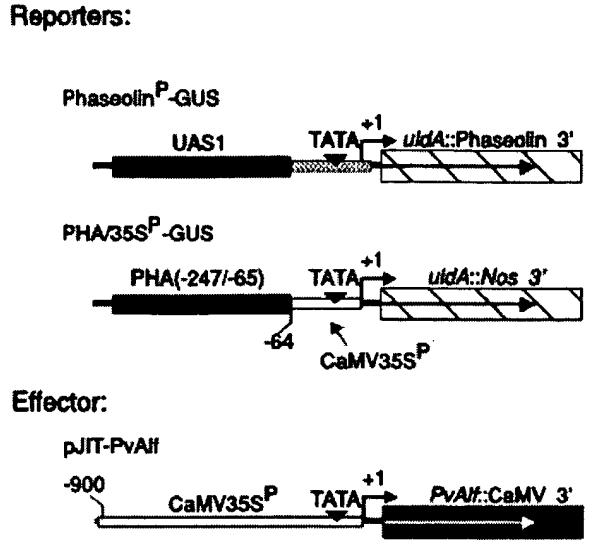

Figure 4. Constructs used in PvAlf *trans*-activation assays.
Promoters and upstream DNA regulatory sequences driving the *uidA* gene encoding β-glucuronidase: Phaseolin^P–GUS, contains a seed-specific β-phaseolin promoter (–302 to +20) comprising upstream activating sequence-1 (UAS1, –302 to –106) and TATA region (Bustos *et al.*, 1991); PHA/35S^P-GUS, has the PHA –247 to –65 enhancer (Riggs *et al.*, 1989) fused to a truncated CaMV 35S promoter (–64) Both genes have β-phaseolin polyadenylation and 3'-flanking sequences (Phaseolin 3') The effector plasmid pJIT–PvAlf contains a PvAlf cDNA inserted in the polylinker region of pJIT82 between CaMV 35S promoter (–900 to +1) and polyadenylation (CaMV 3') sequences.

Gene PvAlf activates seed maturation-specific promoters in bean

A promoter *trans*-activation assay based on particle bombardment of bean cotyledon tissues with recombinant DNA was used to explore the relationship between PvAlf and phaseolin or PHA-L expression. Figure 4 depicts the structures of two reporter constructs designed to monitor transient gene expression: Phaseolin^P–GUS consisted of a β-phaseolin gene fragment (–302 to +20) defined previously as a minimal seed-specific promoter in transgenic tobacco plants (Bustos *et al.*, 1991), driving the *uidA* gene from *Escherichia coli* that encodes β-glucuronidase (GUS); PHA/35S^P–GUS, consisted of an upstream PHA-L promoter fragment (–247 to –65) fused to a CaMV 35S TATA-containing fragment (–64 to +1) also driving the *uidA* reporter gene. The –247 to –65 PHA–L fragment is necessary for seed-specific expression in tobacco (Riggs *et al.*, 1989). The PvAlf effector plasmid pJIT–Alf contained a full-length PvAlf cDNA (encoding amino acids 1–753) under the control of a CaMV35S promoter (–900 to +1) and 35S termination signals (CaMV-3') of plasmid pJIT82 (a gift from Dr D. Helinski, UCSD, La Jolla, CA). To control for bombardment efficiency parallel experiments were carried out with construct pJIT–GUS containing the *uidA* gene in vector pJIT82.

Reporter and effector plasmids were introduced into

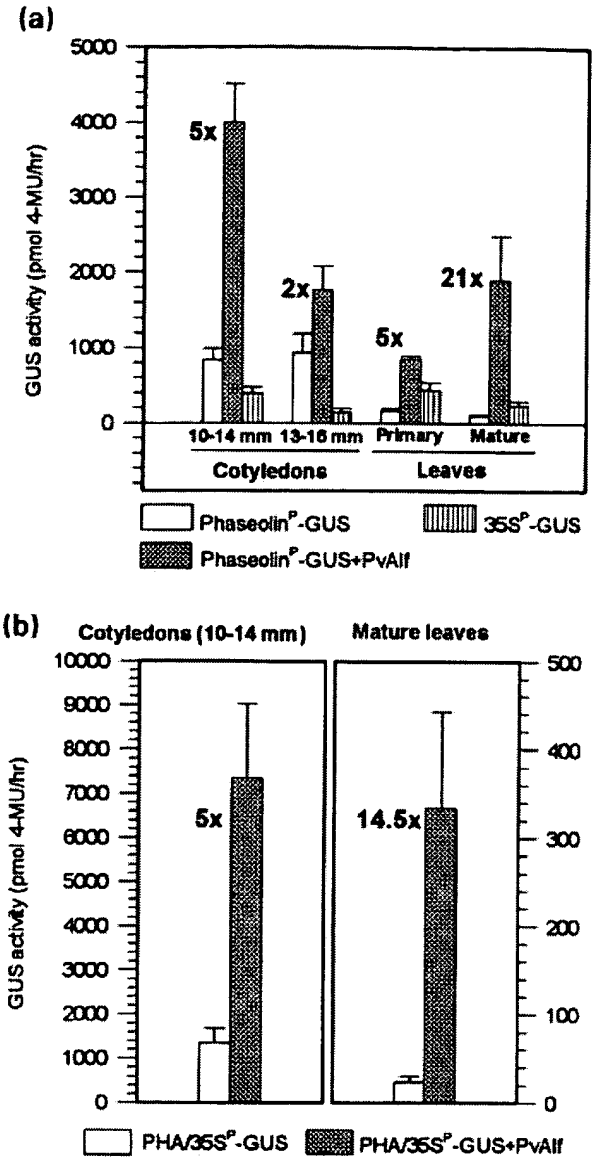

Figure 5. PvAlf activates expression from promoters containing phaseolin and PHA regulatory sequences in cotyledon and leaf tissues.
Reporter and effector plasmids were introduced into bean cotyledon and leaf cells by microprojectile bombardment using a Helium-driven apparatus (Bio-Rad). Bars represent the mean of five to seven individual bombardments ± SE.
(a) Phaseolin reporter gene. Plasmid Phaseolin^P–GUS was bombarded together with an equal amount of either pJIT82 (empty bars) or the effector pJIT–PvAlf (cross-hatched bars). To estimate transfection efficiency, pJIT–GUS was bombarded with and equal amount of pJIT82 (vertical hatch.
(b) PHA reporter gene. Conditions were as in (a), except that plasmid PHA/35S^P–GUS was used as reporter.

cotyledon cells using a Helium-driven Biolistic particle delivery apparatus (Bio-Rad). Figure 5(a) presents data on *trans*-activation of the Phaseolin^P–GUS reporter in cotyledons at two different stages, 10–14 and 13–16 mm, corresponding to age ranges of 14–22 DAF (median 18 DAF), and 21–25 DAF (median 23 DAF). The first cotyledon

stage was chosen to match the period of maximal abundance of endogenous phaseolin and PHA mRNAs, while the later stage corresponded to the period of decline in both types of mRNA. In every case, GUS activities are expressed in picomols (pmol) of 4-MU produced per hour and were calculated after subtracting background values from identical tissue samples that had not been subjected to particle bombardment. All bars represent the average of five to seven repetitions. A 1:1 ratio of effector to reporter plasmid DNA produced significant and consistent enhancements (indicated in boldface type) in reporter gene activity. The effect was more pronounced (fivefold versus twofold) in 10–14 mm cotyledons, although this difference may be the result of higher activity of the CaMV 35S promoter in younger embryos. To determine whether PvAlf activation of the phaseolin promoter requires the presence of other seed-specific factors, the same experiment was performed on primary leaves (3–5 days into germination) and mature leaves. In both cases, a large net activation was observed in the presence of PvAlf although the effect was relatively stronger (21-fold versus fivefold activation) in mature leaves.

Similar experiments were performed with the PHA reporter gene on 11–14 mm cotyledons and mature leaves (Figure 5b). The results were nearly identical to those obtained with the phaseolin driven reporter indicating that both promoters are targets for activation by PvAlf. Moreover, control experiments showed that PvAlf has no effect on a minimal (–64 to +20) phaseolin promoter fragment or on the CaMV 35S promoter (data not shown) indicating that the effect requires specific *cis*-acting elements present in the upstream enhancers of phaseolin and PHA genes. From these experiments we conclude that transient gene expression of reporter genes driven by maturation-specific promoters can be activated by recombinant PvAlf. Moreover, if additional factors are required for PvAlf-mediated activation, they must be present in leaves as well as embryos.

The PvAlf protein contains a transcriptional activation domain near its N-terminus

Although recombinant PvAlf appeared to function as a positive regulator of maturation-specific promoters, it was not immediately obvious whether it acted at the level of transcription or as a regulator (e.g. protein kinase) of endogenous transcription factors. McCarty *et al.* (1991) demonstrated the presence of a transcriptional activation domain near the N-terminus of Vp1; however, the homology of the N-terminal regions of PvAlf and Vp1 is so low (see alignment on Figure 1) that it would be unwise to assume that PvAlf is also active in transcription. Therefore, we ascertained whether tethering PvAlf to an upstream location in a heterologous promoter would lead to transcriptional activation *in vivo*. To that end, three PvAlf ORFs coding for N-terminal polypeptides of increasing length (117, 243 and 756 amino acids) were fused to the DNA-binding domain of the yeast transcription factor GAL4 (amino acids 1–147, Ma and Ptashne, 1987). The GAL4 DNA-binding domain (GAL4DB) recognizes a specific, 17 bp (17-mer) consensus sequence but is insufficient to activate transcription of its own. Figure 6(a) depicts these gene fusions along with a complete GAL4 gene. These plasmids were used to transform yeast strain SFY526 harboring a chromosomal copy of the GAL4-responsive gene GAL1: *lacZ* (Bartel *et al.*, 1993). Transformants were first tested for induction of β-galactosidase activity using a filter assay. Individual colonies were subsequently grown in liquid culture and β-galactosidase activity was quantified by a colorimetric assay (Miller, 1972). Figure 6(b) shows a plot of the amount of β-galactosidase activity (expressed in Miller units) produced by each fusion gene. As expected, the GAL4 DNA-binding domain alone was unable to activate *lacZ* expression. Addition of the first 117 amino acids of PvAlf increased expression by 10- to 15-fold over the

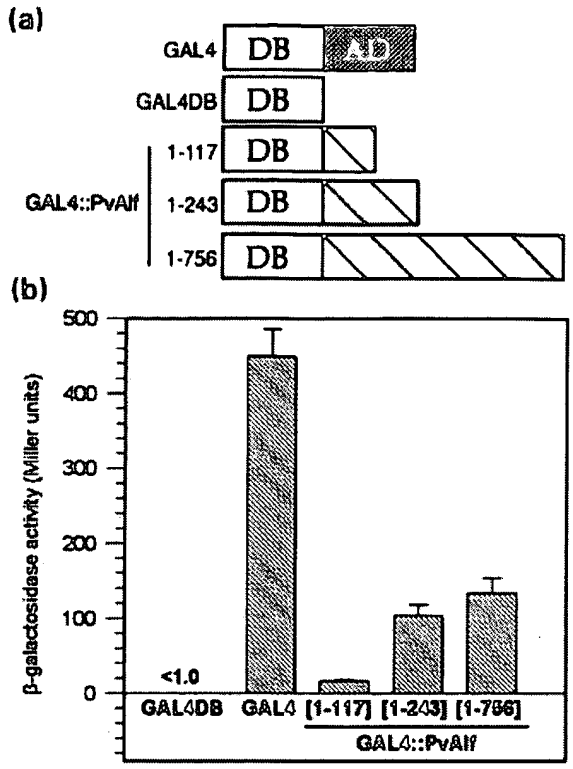

Figure 6. N-terminal fragments of PvAlf function as transcriptional activation domains in yeast.
(a) Diagrams represent fusions between the DNA-binding domain (amino acids 1–147) of GAL4 and PvAlf polypeptides. DB, DNA-binding domain; AD, GAL4 transcriptional activation domain. Cross-hatched boxes indicate PvAlf sequences. Numbers on the left show the length of PvAlf peptides (in amino acids) included in each fusion.
(b) β-galactosidase activity (Miller units) in yeast cultures transformed with GAL4:PvAlf expression plasmids.

GAL4DB value. Although low, this degree of activation was highly reproducible. Adding PvAlf amino acids 118–243 increased activation by another order of magnitude to approximately 140 times the GAL4DB value. Inclusion of remaining amino acids at the C-terminus of PvAlf yielded a very small and more variable increase in β-galactosidase activity. These experiments demonstrated that PvAlf activates transcription when targeted to a promoter upstream location, and that the N-terminal 243 amino acids of PvAlf function as an efficient transcriptional activation domain in yeast.

The function of the PvAlf N-terminal activation domain was confirmed in bean cotyledon cells using the particle bombardment procedure. For that purpose, a reporter gene was used that contained a synthetic, GAL4-responsive promoter driving the gene for chloramphenicol acetyl transferase (CAT). This construct, a generous gift from Dr Jun Ma, is designated as $GAL4_{(17mers)}$:TATA–CAT in Figure 7; it contains 10 copies of a 17 bp consensus GAL4-binding site fused to the CaMV 35S TATA fragment (Ma and Ptashne, 1987). A version of this promoter lacking GAL4-binding sites (TATA–CAT) was used as a negative control. Plant effector plasmids GAL4::$PvAlf_{(1-117)}$ and GAL4::$PvAlf_{(1-243)}$ were based on the same vector (pJIT82) used to demonstrate *trans*-activation by recombinant PvAlf protein (Figure 4). After a 24–36 h period of incubation, the tissue was homogenized and CAT activity quantified using [^{14}C]chloramphenicol and acetyl-CoA (Fromm *et al.*, 1987). Autoradiographs of typical TLC separations are shown in Figure 7(a) The amount of radioactivity corresponding to chloramphenicol and acetylated chloramphenicol was quantified using a Phosphorimager (Molecular Dynamics). Corresponding CAT activity values are displayed in Figure 7(b). A small increase in CAT expression (4-fold) was observed with the GAL4–$PvAlf_{(1-117)}$ construct. By contrast, construct GAL4–$PvAlf_{(1-243)}$ caused a very large increase (71-fold) in the activity of the reporter gene. Both effects were observed only with the construct containing GAL4-binding sites, demonstrating that binding of the GAL4::PvAlf fusion proteins to the promoter was a prerequisite for gene activation. These results were entirely consistent with those obtained previously in yeast and confirmed the function of the PvAlf N-terminal region as a transcriptional activation domain.

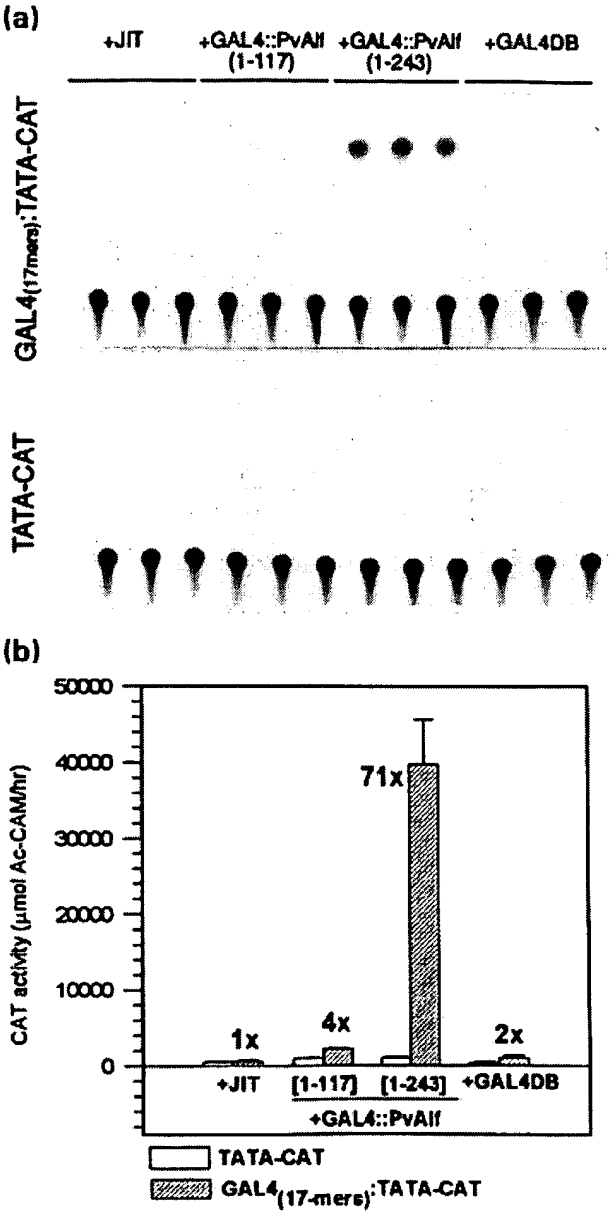

Figure 7. The $PvAlf_{(1-243)}$ N-terminal activation domain is functional in bean cotyledon cells.
Analysis of chloramphenicol acetyl transferase (CAT) activity in immature bean cotyledons co-transfected with CAT reporter genes $GAL4_{(17\text{-mers})}$:TATA–CAT or TATA–CAT and plasmids pJIT82 (+JIT), GAL4::$PvAlf_{(1-117)}$, GAL4::$PvAlf_{(1-243)}$ and GAL4 DNA-binding domain (GAL4DB).
(a) Autoradiographs of TLC separations of [^{14}C]Cam and [^{14}C]Ac–Cam.
(b) Plot of CAT activity (pmol of Ac-Cam produced per h) calculated from TLC chromatographs using a Phosphorimager (Molecular Dynamics). Bars correspond to the mean ± SE. Numbers in boldface type indicate fold enhancement over the control (+JIT).

The PvAlf N-terminal transcriptional activation region resembles acidic domains of eukaryotic transcription factors GCN4, HAP4 and VP16

Figure 8 compares the amino acid sequences of the activation domains of plant VAlf proteins PvAlf and Vp1, with analogous activation domains of yeast factors HAP4 (Forsburg and Guarente, 1989) and GCN4 (Hope and Struhl, 1986) and the herpes simplex virus (HSV) virion protein VP16 (Regier *et al.*, 1993). The corresponding N-terminal region of *A. thaliana* ABI3 (AtABI3) and a mutant VP16 protein, VP16ΔNC413–429, that retains most of the transcriptional activation capacity of VP16 (Triezenberg *et al.*, 1988) are also shown. All seven proteins share a central subdomain highly enriched in serine, threonine and proline

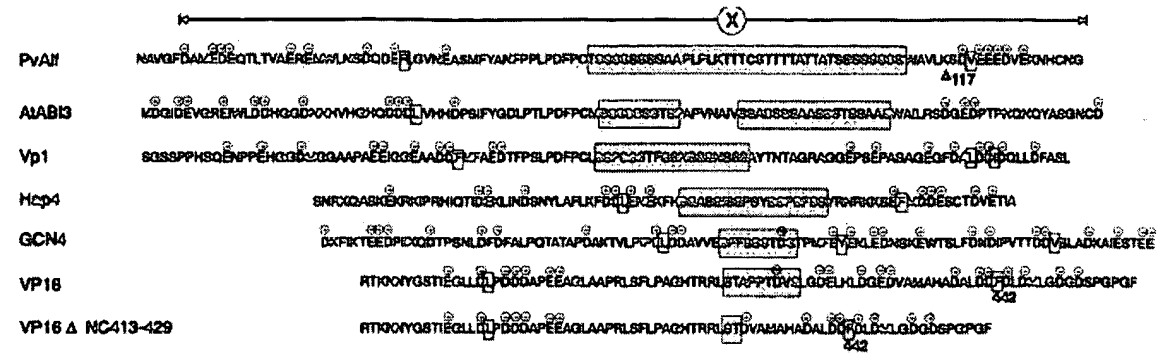

Figure 8. The acidic activation domains of plant VAlf proteins.
The amino acid sequence of the acidic activation domains of PvAlf and Vp1 are compared with a corresponding region in domain I of ABI3 (AtABI3), with the acidic domains of yeast transcription factors HAP4 and GCN4, and with the C-terminal acidic domain of VP16. VP16ΔNC413–429 is a mutant VP16 product lacking amino acids 413–429 that activates transcription at nearly wild-type levels. Acidic residues (glutamic and aspartic acid) are marked with circled '–' signs. Clusters of serine, threonine and proline residues (STP clusters) are highlighted with stippled boxes. Bulky hydrophobic amino acids in the acidic subdomains are shown enclosed in rectangles.

(STP clusters, stippled boxes) surrounded by two acidic (negatively charged) regions. The sequences shown in Figure 8 have been arranged in order of decreasing length of the central subdomain which varies from approximately 60 amino acids in PvAlf to only 17 in GCN4. The boundaries of this central region are delimited by the two acidic subdomains. Secondary structure prediction by the Chou and Fassman method (Chou and Fassman, 1974) suggests that the STP clusters confer upon the central domain a high propensity to form β-turns. The importance of the acidic subdomains rich in glutamic and aspartic acid can be deduced from their similarity to acidic activation domains of GCN4 (Hope and Struhl, 1986) and VP16 (Cress and Triezenberg, 1991). In these factors, bulky hydrophobic amino acids (phenylalanine, tyrosine and valine) of the acidic subdomains are critical for transcriptional activation. Similar hydrophobic residues are also present in the acidic subdomains of plant VAlfs (square symbols). Although initially (Hope and Struhl, 1986) the activation domain of GCN4 was localized to the acidic region on the C-terminal side of the central STP domain (Figure 8), recent work has demonstrated that GCN4 contains a second acidic subdomain on the N-terminal side of the STP cluster that is also active in transcription (Drysdale *et al.*, 1995). In the case of PvAlf, a 10-fold increase in reporter gene expression resulted from adding the 1–117 peptide that contains the N-terminal acidic subdomain and the STP cluster. Approximately the same 10-fold increase resulted from adding the C-terminal acidic subdomain included in the 118–243 peptide. This is consistent with the presence of two acidic subdomains on either side of the STP cluster contributing to the overall activity of the PvAlf activation domain. By contrast, the significance of the central STP cluster remains unexplored in plants and yeast.

Discussion

Gene *PvAlf* encodes a large, 756 amino acid protein similar to the late embryogenesis regulatory factors ABI3 and Vp1 of *Arabidopsis* and maize; together with OsVp1 (the rice equivalent of Vp1) these four proteins compose a class, designated here as VAlf. A computer assisted alignment of VAlf primary amino acid sequences (Figure 1) extends the observation made first by Giraudat *et al.* (1992) and later by Hattori *et al.* (1994), that the similarity among these proteins is confined to four domains (I–IV). The conservation of VAlf domains I–IV contrasts with the widely divergent sequences of intervening segments that separate those domains. Such alternation of high sequence similarity and divergence constitutes strong evidence that VAlf proteins consist of several structural modules. A modular organization is common among transcription factors whose various functions such as DNA-binding, activation, protein:protein contacts, and small ligand binding are segregated into distinct and largely independent globular domains. At the same time, regions of low sequence similarity could represent functional specializations reflecting the different cellular environments in which each VAlf protein must perform. New procedures for legume transformation now offer a unique opportunity to address these questions by comparing the functions of hybrid ABI3–PvAlf proteins in transgenic *Arabidopsis* and bean (or soybean) plants. The widespread occurrence of VAlf proteins also suggests that they are essential for the normal development of plant species exhibiting a maturation phase of embryogeny, and raises the question of whether species that do not undergo embryonic maturation, such as ferns, lack functional VAlf genes.

Homozygous *abi3* and *vp1* seeds fail to accumulate maturation-specific products in endosperm and embryonic

cells (McCarty *et al.*, 1991; Nambara *et al.*, 1992, 1994; Parcy *et al.*, 1994) and evidence that the products of these two loci are active in transcription continues to mount. For instance, recombinant Vp1 and OsVp1 proteins transiently expressed in electroporated cell protoplasts enhance the activity of maize Em and C1 promoters (Hattori *et al.*, 1992, 1994; McCarty *et al.*, 1991), and ectopic expression of ABI3 in transgenic seedlings leads to activation of maturation and late embryogenesis markers in leaves. We found that a recombinant *PvAlf* gene also activates gene expression from the promoters of two bean maturation-specific markers, phaseolin and PHA-L, in embryonic and non-embryonic cell types. Preliminary results suggest that the PvAlf effector plasmid pJIT-Alf may also induce expression of phaseolin mRNA in bean leaf cells. The fact that ABI3 and PvAlf proteins seem to function in leaves is very significant, since neither gene is normally expressed outside of the seed environment (Figure 3 in this paper and Giraudat *et al.*, 1992). This indicates that no other seed-specific factors may be needed for transcriptional activation by PvAlf or ABI3. Alternatively, a factor present in leaves may be able to substitute partially or completely for seed-specific proteins that normally assist ABI3 and PvAlf function in the embryos. For instance, different R genes that regulate anthocyanin biosynthesis in maize are expressed with distinct tissue specificities (e.g. S gene in aleurone versus P gene in anther and coleoptile or Lc gene in midrib, pericarp and other tissues) but encode transcription factors with essentially the same function (Ludwig *et al.*, 1989).

The time course of phaseolin, PHA-L and PvAlf mRNA accumulation in developing bean embryos raises several interesting questions for future investigation. Phaseolin and lectin genes are coordinately regulated between 10 and 27 DAF. Their induction several days after the initial increase in PvAlf mRNA expression is consistent with their position downstream of PvAlf in a control hierarchy. This, and the fact that PvAlf is sufficient to activate *de novo* GUS expression from phaseolin and PHA-L promoters support a role for the PvAlf protein as a positive regulator of maturation-specific gene expression in bean. More definitive proof will come from immunological detection of PvAlf protein in embryonic tissues using specific antibodies, and from the use of an antisense RNA approach to suppress PvAlf synthesis in transgenic embryos. Presumably, PvAlf activates gene transcription by interacting, directly or via a facilitator protein, with specific *cis*-acting motifs of phaseolin and PHA-L promoters. The only obvious similarities between the phaseolin and PHA-L upstream sequences present in reporter constructs Phaseolin^{P}–GUS and PHA/35S^{P}–GUS are three instances of the motif $^{5'}$CATGCAY$^{3'}$, homologous to RY-repeats of many legume seed-expressed genes (Dickinson *et al.*, 1992; Hoffman and Donaldson, 1985; Thomas, 1993), and several occurrences of the sequence $^{5'}$G^{A}/$_{C}$CAC^{G}/$_{C}$TCA$^{3'}$. The latter binds to two basic leucine-zipper (bZIP) proteins, PvSF1 and VBP1, recently cloned from bean embryos in our laboratory (Chern *et al.*, submitted). Phaseolin and PHA-L RY-repeats are similar to the Sph1 element required for Vp1-mediated activation of the C1 promoter in maize cell protoplasts (Hattori *et al.*, 1992). We are currently investigating the possibility that the RY-repeats mediate transcriptional activation by PvAlf. Experiments are also underway to determine any possible interactions of PvAlf with cloned bZIP factors that bind to both promoters, or with other transcription factors and *cis*-acting elements that may also regulate these promoters.

Phaseolin and PHA-L mRNAs decrease in abundance and eventually disappear between 22 and 35 DAF while PvAlf continues to be expressed. A similar phenomenon has been described in *Arabidopsis* (Parcy *et al.*, 1994) where ABI3-dependent, maturation genes such as cruciferin, are downregulated late in embryogenesis although the steady-state level of ABI3 mRNA remains constant. This pattern is also independent of the concentration of endogenous ABA in embryos. These observations could be explained by the presence of negative regulatory factors that repress *MAT* gene expression before abscision and desiccation. Run-off transcription experiments carried out with phaseolin, β-conglycinin and Kunitz trypsin inhibitor (Kti) genes have indicated that repression of maturation-specific genes may occur both at the transcriptional and post-transcriptional levels (Barker *et al.*, 1988; Chappell and Chrispeels, 1986; Jofuku and Goldberg, 1989; Walling *et al.*, 1986). Apparently, ABI3-dependent late embryogenesis abundant (*LEA*) genes (Parcy *et al.*, 1994) somehow remain impervious to this repression and continue to be expressed well into the desiccation stage (Finkelstein, 1993; Hughes and Galau, 1991).

The functional importance of acidic, amphipathic α-helices within the Vp1 transcriptional activation domain has been suggested by McCarty *et al.* (1991) based on their similarity with acidic sequences in other eukaryotic transcription factors (Giniger and Ptashne, 1987). More recently, the importance of α-helices in these domains has been called into question (Cress and Triezenberg, 1991; Van Hoy *et al.*, 1993), and now it appears that certain bulky hydrophobic residues embedded within a negatively charged structure are what dictates their activity. A very conspicuous feature of all four plant VAlf activation domains are the large serine–threonine–proline rich (STP) clusters that separate the acidic subdomains (Figure 8). These clusters resemble degradation sequences of plasma membrane proteins whose half-life is regulated by O-linked glycosylation (Kozarski *et al.*, 1988). O-linked GlucNac side-chains have been demonstrated on many nuclear proteins (reviewed by Hart *et al.*, 1989) and near the activation regions of animal transcription factors (Jackson and Tjian, 1988). Another attractive possibility is that the STP clusters

represent targets for serine kinases similar to glycogen synthase kinase-3 of animal cells (Kemp and Pearson, 1990). The ABA-insensitive-1 (*ABI1*) gene of *Arabidopsis* encodes a novel type of Ca^{+2}-regulated phosphatase 2C (Leung *et al.*, 1994; Meyer *et al.*, 1994) indicating that phosphorylation/dephosphorylation reactions play a major role in the ABA signal transduction pathway (Bowler and Chua, 1994; Rock and Quatrano, 1994). We speculate that the STP clusters may have a regulatory role either by controlling the half-life of VAlf activation domains, or as phosphorylation sites. The PvAlf→ref phaseolin and PvAlf→PHA-L regulatory systems described here will allow a critical evaluation of these and other hypothetical functions of VAlfs in dicots.

Experimental procedures

Cloning of PvAlf

A rapid amplification of cDNA ends (3'-RACE, Frohman *et al.*, 1988) strategy was used to clone PvAlf. Total RNA was obtained from maturing, 5–10 mm long embryos of the common bean (cv. 'Tendergreen') by extraction with guanidinium thiosulfate (Sambrook *et al.*, 1989). Poly(A)$^+$ RNA was isolated using the PolyATtract™ magnetic-bead system (Promega) and cDNA was synthesized with Superscript™ Reverse Transcriptase (Gibco BRL). This cDNA served as template for all PCR amplifications with the exception of those for the 5'-RACE procedure. Two degenerate oligonucleotides were synthesized, prALF3'R-1 (5'-ATGGARGAYATHGGNAC-3'; 48 permutations) and prALF3'R-2 (5'-GTNTGGAAYATGMGNTA-3'; 64 permutations). A first amplification was carried out with one gene-specific primer (prALF3'R-1) and the 3' adapter primer. After 35 cycles, remaining primers were eliminated by ultrafiltration, diluted and a second amplification was performed with prALF3'R-2 and 3' adapter. The products of the second amplification were cloned (TA cloning™, Invitrogen) and sequenced. Two nested, specific downstream primers were synthesized based on the 3'-RACE product sequence, prALFDWN-1 (5'-GGTTTCACACCTTGTTG-3') and prALFDWN-2 (5'-GCTGG GTTTTCTGCGAT-3'). A degenerate oligonucleotide, prALFUP-1, (5'-CTYCCNGAYTTYCCNTG-3'; 128 permutations) was synthesized from a conserved region (LPDFP) near the N-terminus of Vp1-like proteins. Nested PCR reactions were performed with the degenerate primer prALFUP-1 and either of the two PvAlf-specific primers, prALFDWN-1 and prALFDWN-2. A major product of the second reaction, 1.8 kb in length, was purified by agarose electrophoresis, cloned (TA cloning™) and sequenced to confirm its identity. A new primer was synthesized, prALF5'R (5'-AGGATCGAAGAAATCATTGGC-3') and used in the 5'-RACE System (GIBCO-BRL) to amplify the 5'-end of PvAlf mRNA using Poly(A)$^+$ RNA as template. The products of the 5' RACE reaction were analyzed by Southern blot hybridization using the 1.8 kb PvAlf fragment. The largest cross-hybridizing products were isolated by gel electrophoresis and cloned. Recombinant clones were analyzed by colony hybridization (Sambrook *et al.*, 1989) with the modification that colonies were poked through a nitrocellulose filter into antibiotic-containing LB agar plates and the filter/plate assembly incubated together. Inserts from hybridizing colonies were sequenced to identify those containing the beginning of the PvAlf open reading frame.

Based on the collected sequences of 3' and 5'-RACE products, oligonucleotide primers prAlf5' (5'-CCGTCGACGCAAAGATGGAGTGTGAAGTGAAG-3') and prAlf3' (5'-CCGATATCCTGTTGACAGCCTCCATTGC-3') were synthesized to permit PCR amplification of the entire predicted coding region of PvAlf from cDNA. Two independent PCR reactions were performed and their products cloned. One of these full-length clones was completely sequenced to verify the sequence of PvAlf.

RNA expression analysis and DNA hybridization

Various tissues of *Phaseolus vulgaris* were frozen in liquid nitrogen and total cellular RNA was extracted by a hot phenol method (Meier *et al.*, 1993). The embryos were separated from seed coats and endosperm before freezing. RNAs were denatured with formaldehyde, resolved by agarose gel electrophoresis and transferred to nylon membranes (Nytran™, S&S). The membranes were pretreated in hybridization buffer (1 M NaCl, 10% dextran sulfate, 1% SDS, 150 μg denatured salmon sperm DNA) for 2 hr. at 65°C and hybridized in the same solution to a 1.8 kb PvAlf probe labeled with ^{32}P-dCTP by the random primed method (USB). After washing (twice 30 min in 2× SSC, 1% SDS followed by twice 30 min in 0.2× SSC, 1% SDS at 60°C), the blots were autoradiographed using Kodak X-Omat X-ray film. Blots were stripped, and re-hybridized to similarly prepared probes for phaseolin, PHA-L and 18S rRNA. Total genomic DNA was isolated from etiolated bean seedlings by the CTAB method (Taylor and Powell, 1982), separated by electrophoresis on a 0.75% agarose/TAE gel, immobilized on Nytran™ (S&S) filters by capillary transfer and probed with an internal *Eco*RI restriction fragment of PvAlf as above.

Plasmid constructs

Hybrid genes containing PvAlf full-length or C-terminal truncated fragments were fused in-frame with the GAL4 DNA-binding domain in the yeast expression vector pGBT9 (Clontech). In every case, PvAlf sequences were amplified by PCR from a full-length cDNA clone (pPvAlf) using synthetic primers containing restriction sites. The nucleotide sequences of each primer were as follows:

PvAlf(1–756): Alf3'-*Sma*I (TGGCCCGGGGATGGAGTGTGAAGTGAAG) and Alf5'-*Sal*I (GGGGTCGACTAGTTTCGGTGCGATGAC);

PvAlf(1–117): Alf1-*Eco*RI (GGGGAATTCATGGAGTGTGAAGTGAAG) and Alf2-*Bam*HI (GGGGGATCCCTTCAAGACGGCCCGTG);

PvAlf(1–243): Alf5'-*Sal*I (GGGGTCGACTAGTTTCGGTGCGATGAC) and Alf6-*Sal*I (GGGGTCGACCTCTCCTTTGATAACTTGAC).

All PCR products were first cloned in the TA™ cloning vector pCR2000 (Invitrogen), then excised with appropriate restriction enzymes and ligated into pGBT9. For functional analysis of PvAlf fragments in bean the GAL4:PvAlf$_{(1-117)}$ and GAL4:PvAlf$_{(1-243)}$ fusions were excised from pGBT9 by digestion with *Hin*dIII and *Bam*HI or *Hin*dIII and *Sal*I, and subcloned downstream of the CaMV 35S promoter in vector pJIT82 (a gift from Professor Donald Helinski, Department of Biology, University of California, San Diego, La Jolla, CA 92083, USA). The control plasmid pGAL4DB:JIT coding for the GAL4 DNA-binding domain was made by excising the PvAlf fragment of pGAL4:PvAlf$_{(1-117)}$/JIT with *Eco*RI and religating the remaining plasmid with T4 DNA ligase.

Yeast transformation and β-galactosidase assays

Functional assays were conducted in yeast strain SFY526 (Bartel *et al.*, 1993) that has a *lacZ* reporter gene driven by a GAL1

promoter and transformation markers *trp1-901* and *leu2-3*. Yeast cells were transformed with expression vectors carrying different GAL4:PvALF fusions by the Li-acetate method as described by Ito *et al.* (1983) and modified by Schiestl and Gietz (1989), Hill *et al.* (1991), and Gietz *et al.* (1992). Quantitative β-galactosidase assays were performed by growing individual colonies in minimal medium containing 2% dextrose to mid-log phase (OD_{600} ~ 1.0). Cell culture (0.1 ml) was added to 0.7 ml Z-buffer (60 mM $Na_2HPO_4.7H_2O$, 40 mM $NaH_2PO_4.H_2O$, 10 mM KCl, 1 mM $MgSO_4.7H_2O$ at pH 7.0 and 40 mM β-mercaptoethanol) and the cells lysed by the addition of 50 µl $CHCl_3$ and 50 µl 0.1% SDS. After the addition of 0.16 ml *o*-nitrophenylgalactoside solution (4 mg ml^{-1} in 0.1 M phosphate buffer at pH 7.0), the samples were incubated for 1 h at 37°C. Reactions were stopped with 0.4 ml Na_2CO_3 and cell debris removed by centrifugation at 13 000 *g* for 10 min. The absorbance at 420 nm was determined and β-galactosidase activity expressed as in Miller (1972). Filter assays for qualitative β-galactosidase activity detection were performed by placing a sterile Whatman #1 filter on top of agar plates containing the transformed colonies which were subsequently submerged for 10 sec in liquid nitrogen to lyse the cells. The filters were placed on top of Whatman filters that had been presoaked in 1.8 ml Z-buffer containing X-gal at 0.33 mg ml^{-1}. The filters were incubated at 30°C until blue color developed.

Particle bombardment of bean cotyledons and measurement of β-glucuronidase (GUS) and chloramphenicol acetyl transferase (CAT) activity

Cotyledons from immature bean seeds (~ 17 days after flowering) were sliced longitudinally and placed on solid Gamborgós G5 medium containing 0.8% agar, 3% sucrose and 0.75 M mannitol. Leaf discs (1 inch in diameter) were cut from mature bean leaves taking care to avoid major veins and placed on the same medium. Tungsten (1.7 µm ϕ) or gold (1.6 µm ϕ) microcarriers (0.2 mg per bombardment) were co-precipitated with equal amounts (0.5 µg per bombardment) of reporter and effector or reporter and vector (pJIT82) plasmids. Microcarriers were spotted on to macrocarriers and delivered on to the surface of the tissues with a He-driven particle gun (BioRad) at 1550 psi. Bombarded tissues were incubated for 1–2 days at 28°C in the dark and then ground with a mortar and pestle in 400 µl of GUS extraction buffer or 0.25 M Tris–HCl at pH 7.5 (for CAT assays). Cell debris was spun down for 5 min at 13 000 *g*. GUS activity was measured as described in Jefferson (1987) and CAT activity as in Fromm *et al.* (1987).

Acknowledgments

The authors thank Dr Jun Ma for plasmids pMa556 and pMa558, Dr Donald Helinski for plasmid pJIT82, M.-S. Chern for technical advice during cloning and Dr Carol Greitner for expert supervision of plant stocks. This research was supported by grants from the National Science Foundation (MCB-9219203) and the USDA NRI-CGP (USDA-9303090) to M. M. Bustos, and by a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft to H. G. Eiben. A. J. Bobb was the recipient of a predoctoral Research Assistantship from the Maryland Agricultural Experimental Station (MAES).

References

Barker, S.J., Harada, J.J. and Goldberg, R.B. (1988) Cellular localization of soybean storage protein mRNA in transformed tobacco plants. *Proc. Natl Acad. Sci. USA*, **85**, 458–462.

Bartel, P.L., Chien, C.-T., Sternglanz, R. and Fields, S. (1993) Elimination of false positives that arise in using the two-hybrid system. *BioTechniques*, **14**, 920–924.

Bowler, C. and Chua, N.-H. (1994) Emerging themes of plant signal transduction. *Plant Cell*, **6**, 1529–1541.

Bustos, M.M., Begum, D., Kalkan, F.A., Battraw, M.J. and Hall, T.C. (1991) Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. *EMBO J.* **10**, 1469–1479.

Casey, R., Domoney, C. and Ellis, N. (1986) Legume storage proteins and their genes. In *Oxford Surveys of Plant Molecular and Cell Biology*, Volume 3 (Miflin, B.J. ed.). Oxford: Oxford University Press, pp. 1–95.

Chappell, J. and Chrispeels, M.J. (1986) Transcriptional and post-transcriptional control of phaseolin and phytohemagglutinin gene expression in developing cotyledons of *Phaseolus vulgaris*. *Plant Physiol.* **81**, 50–54.

Chou, P.Y. and Fasman, G.D. (1974) Prediction of protein conformation. *Biochemistry*, **13**, 222–245.

Cress, W.D. and Triezenberg, S.J. (1991) Critical structural elements of the VP16 transcriptional activation domain. *Science*, **251**, 87–90.

Crouch, M.L. (1987) Regulation of gene expression during seed development in flowering plants. In *Developmental Biology: A Comprehensive Synthesis*, Volume 5 (Browder, L.W. ed.). New York: Plenum Press, pp. 367–404.

Dickinson, C.D., Evans, R.P. and Nielsen, N.C. (1992) RY repeats are conserved in the 5'-flanking region of legume seed protein genes. *Nucl. Acids Res.* **16**, 371.

Drysdale, C.M., Dueñas, E., Jackson, B.M., Reusser, U., Braus, G.H. and Hinnebusch, A.G. (1995) Transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Mol. Cell. Biol.* **15**, 1220–1233.

Finkelstein, R.R. (1993) Abscissic acid-insensitive mutations provide evidence for stage-specific signal pathways regulating expression of an *Arabidopsis* late embryogenesis-abundant (*lea*) gene. *Mol. Gen. Genet.* **238**, 401–408.

Finkelstein, R.R. and Sommerville, C. (1990) Three classes of abscisic acid (ABA)-insensitive mutations of *Arabidopsis* define genes that control overlapping subsets of ABA responses. *Plant Physiol.* **94**, 1172–1179.

Forsburg, S.L. and Guarente, L. (1989) Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. *Genes Devel.* **3**, 1166–1178.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl Acad. Sci. USA*, **85**, 8998–9002.

Fromm, M., Callis, J., Taylor, L.P. and Walbot, V. (1987) Electroporation of DNA and RNA into plant protoplasts. *Methods Enzymol.* **153**, 351–366.

Galau, G.A., Jakobsen, K.S. and Hughes, D.W. (1991) The controls of late dicot embryogenesis and early germination. *Physiol. Plant.* **81**, 280–288.

Gietz, D., St. Jean, A., Woods, R.A. and Schiestl, R.H. (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucl. Acids Res.* **20**, 1425–1425.

Giniger, E. and Ptashne, M. (1987) Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit. *Nature*, **330**, 670–673.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J. and Parcy, F. (1992) Isolation of the *Arabidopsis ABI3* gene by positional cloning. *Plant Cell*, **4**, 1251–1261.

Hart, G.W., Haltiwanger, R.S., Holt, G.D. and Kelly, W.G. (1989) Glycosylation in the nucleus and cytoplasm. *Ann. Rev. Biochem.* **58**, 841–874.

Hattori, T., Vasil, V., Rosenkrans, L., Hannah, L.C. and McCarty, D.R. (1992) The *Viviparous*-1 gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Devel.* **6**, 609–618

Hattori, T., Terada, T. and Hamasuma, S.T. (1994) Sequence and functional analysis of the rice gene homologous to the maize *VP1*. *Plant Mol. Biol.* **24**, 805–810

Hill, J., Donald, K.A. and Griffiths, D.E. (1991) DMSO-enhanced whole cell yeast transformation. *Nucl. Acids Res.* **19**, 5791–5791.

Hoffman, L.M. and Donaldson, D.D. (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. *EMBO J.* **4**, 883–889.

Hope, I.A. and Struhl, K. (1986) Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell*, **46**, 885–894.

Hughes, D.W. and Galau, G.A. (1991) Developmental and environmental induction of Lea and LeaA messenger RNAs and the post-abscission program during embryo culture. *Plant Cell*, **3**, 605–618.

Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**,163–168.

Jackson, S.P. and Tjian, R. (1988) O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. *Cell*, **55**, 125–133.

Jefferson, R.A. (1987) Assaying chimeric genes in plants: The GUS gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387–405.

Jofuku, K.D. and Goldberg, R.B. (1989) Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. *Plant Cell*, **1**, 1079–1093.

Kemp, B.E. and Pearson, R.B. (1990) Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* **15**, 342–346.

Koornneef, M., Jorna, M.L., Brinkhorst-van der Swan, D.L.C. and Karssen, C.M. (1982) The isolation of abscisic acid (ABA) deficient mutants by selection of induced revertants in non-germinating gibberellin sensitive lines of *Arabidopsis thaliana* (L.) Heynh. *Theor. Appl. Genet.* **61**, 385–393.

Koornneef, M., Reuling, G. and Karssen, C.M. (1984) The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. *Physiol. Plant.* **61**, 377–383.

Koornneef, M., Hanhart, C.J., Hilhorst, H.W.M. and Karssen, C.M. (1989) *In vivo* inhibition of seed development and reserve protein accumulation in recombinants of abscisic acid biosynthesis and responsiveness mutants in *Arabidopsis thaliana*. *Plant Physiol.* **90**, 463–469.

Kozarski, K., Kingsley, D. and Krieger, M. (1988) Use of a mutant cell line to study the kinetics and function of O-linked glycosylation of low density lipoprotein receptor. *Proc. Natl Acad. Sci. USA*, **85**, 4335–4339.

Kriz, A.R., Wallace, M.S. and Paiva, R. (1990) Globulin gene expression in embryos of maize *viviparous* mutants. *Plant Physiol.* **92**, 538–542.

Leung, J., Bouvier-Durand, M., Morris, P.-C., Guerrier, D., Chefdor, F. and Giraudat, J. (1994) *Arabidopsis* ABA response gene *ABI1*: Features of a calcium-modulated protein phosphatase. *Science*, **264**, 1448–1452.

Long, S.R., Dale, R.M.K. and Sussex, I.M. (1981) Maturation and germination of *Phaseolus vulgaris* embryonic axes in culture. *Planta*, **153**, 405–415.

Ludwig, S.R., Habera, L.F., Dellaporta, S.L. and Wessler, S.R. (1989) *Lc*, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the *myc*-homology region. *Proc. Natl Acad. Sci. USA*, **86**, 7092–7096.

Ma, J. and Ptashne, M. (1987) Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell*, **48**, 847–853.

McCarty, D.R., Carson, C.B., Lazar, M. and Simonds, S.C. (1989) Transposable element-induced mutations of the *Viviparous*-1 gene in maize. *Devel. Gen.* **10**, 473–481.

McCarty, D.R., Hattori, T., Carson, C.B., Vasil, V., Lazar, M. and Vasil, I.K. (1991) The *Viviparous*-1 developmental gene of maize encodes a novel transcriptional activator. *Cell*, **66**, 895–905.

Meier, B.M., Shaw, N. and Slusarenko, A.J. (1993) Spatial and temporal accumulation of defense gene transcripts in bean (*Phaseolus vulgaris*) leaves in relation to bacteria-induced hypersensitive cell death. *Mol. Plant–Microbe Inter.* **6**, 453–466.

Meyer, K., Leube, M.P. and Grill, E. (1994) A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science*, **264**, 1452–1455.

Miller, J.H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Murray, M.G., and Kennard, W.C. (1984) Altered chromatin conformation of the higher plant gene phaseolin. *Biochemistry*, **23**, 4225– 4232.

Nambara, E., Naito, S. and McCourt, P. (1992) A mutant of *Arabidopsis* which is defective in seed development and storage protein accumulation is a new *abi3* allele. *Plant J.* **2**, 435–441.

Nambara, E., Keith, K., McCourt, P. and Naito, S. (1994) Isolation of an internal deletion mutant of the *Arabidopsis thaliana ABI3* gene. *Plant Cell Physiol.* **35**, 509–513.

Neill, S.J., Horgan, R. and Parry, A.D. (1986) The carotenoid and abscisic acid content of viviparous kernels and seedlings of *Zea mays* L. *Planta*, **169**, 87–96.

Ohara, O., Dorit, R.L. and Gilbert, W. (1989) One-sided polymerase chain reaction: the amplification of cDNA. *Proc. Natl Acad. Sci. USA*, **86**, 5673–5677.

Osborn, T.C., Burow, M. and Bliss, F.A. (1988) Purification and characterization of arcelin seed protein from common bean. *Plant Physiol.* **86**, 399–405.

Paiva, R. and Kriz, A.L. (1994) Effect of abscisic acid on embryo-specific gene expression during normal and precocious germination in normal and *viviparous* maize (*Zea mays*) embryos. *Planta*, **192**, 332–339.

Pang, P.P., Pruitt, R.E. and Meyerowitz, E.M. (1988) Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of *Arabidopsis thaliana*. *Plant Mol. Biol.* **11**, 805–820.

Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M. and Giraudat, J. (1994) Regulation of gene expression programs during Arabidopsis seed development: Roles of the *ABI3* locus and of endogenous abscisic acid. *Plant Cell*, **6**, 1567–1582.

Pla, M., Gomez, J., Goday, A. and Pagès, M. (1991) Regulation of the abscisic acid-responsive gene *rab28* in maize *viviparous* mutants. *Mol. Gen. Genet.* **230**, 394–400.

Regier, J.L., Shen, F. and Triezenberg, S.J. (1993) Pattern of aromatic and hydrophobic amino acids critical for one of two subdomains of the VP16 transcriptional activator. *Proc. Natl Acad. Sci. USA*, **90**, 883–887.

Riggs, C.D., Voelker, T.A. and Chrispeels, M.J. (1989) Cotyledon nuclear proteins bind to DNA fragments harboring regulatory elements of phytohemagglutinin genes. *Plant Cell*, **1**, 609–621.

Robertson, D.S. (1955) The genetics of vivipary in maize. *Genetics*, **40**, 745–760.

Rock, C.D. and Quatrano, R.S. (1994) Insensitivity is in the genes. *Curr. Biol*, **4**, 1013–1015.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schiestl, R.H. and Gietz, R.D. (1989) High efficiency transformation of intact cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16**, 339–346.

Staswick, P. and Chrispeels, M.J. (1984) Expression of lectin genes during seed development in normal and phytohemagglutinin deficient cultivars of *Phaseolus vulgaris*. *J. Mol. Appl. Genet.* **2**, 525–535.

Taylor, B.H. and Powell, A.L.T. (1982) Isolation of plant DNA and RNA. *Focus*, **4**, 4–6.

Thomas, T.L. (1993) Gene expression during plant embryogenesis and germination: An overview. *Plant Cell*, **5**, 1401–1410.

Triezenberg, S.J., Kingsbury, R.C. and McKnight, S.L. (1988) Functional dissection of VP16, the *trans*-activator of herpes simplex virus immediate early gene expression. *Genes Devel.* **2**, 718–729.

Van Hoy, M., Leuther, K.K., Kodadek, T. and Johnston, S.A. (1993) The acidic activation domains of the GCN4 and GAL4 proteins are not α helical but form β sheets. *Cell*, **72**, 587–594.

Walling, L., Drews, G.N. and Goldberg, R.B. (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc. Natl Acad. Sci. USA*, **83**, 2123–2127.

Genbank Data Library accession number U28645 (*P. vulgaris* PvAlf gene).

Plant Molecular Biology **41**: 577–585, 1999.

Mini-review

Multifunctionality and diversity within the plant *MYB*-gene family

Hailing Jin & Cathie Martin
Department of Genetics, John Innes Centre, Colney, Norwich, UK

Received 15 July 1999; accepted in revised form 9 September 1999

Key words: gene evolution, gene redundancy, MYB transcription factors

Abstract

MYB proteins constitute a diverse class of DNA-binding proteins of particular importance in transcriptional regulation in plants. Members are characterised by having a structurally conserved DNA-binding domain, the MYB domain. Different categories of MYB proteins can be identified depending on the number of imperfect repeats of the MYB domain they contain. It is likely that single MYB-domain proteins, a class of expanding importance in plants, bind DNA in a different way than two-repeat or three-repeat MYB proteins, and these groups are therefore likely to have different functions. The two-repeat (R2R3) MYB family is the largest family characterised in plants, and there are estimated to be over 100 members in *Arabidopsis*. Functions of MYB proteins in plants include regulation of secondary metabolism, control of cellular morphogenesis and regulation of meristem formation and the cell cycle. Although functional similarities exist between R2R3 MYB proteins that are closely related structurally, there are significant differences in the ways very similar proteins function in different species and also within the same organism. Therefore, despite the large number of R2R3 MYB proteins in plants, it is unlikely that many are precisely redundant in their functions, but more likely that they share overlapping functions.

Introduction

When the first plant regulatory gene was sequenced, the *C1* gene of maize, it was recognised to encode a transcription factor from its similarity to the relatively well characterised mammalian transcription factor c-MYB (Paz-Ares *et al.*, 1987). Since then the number of proteins with sequence similarity to the MYB domain has increased enormously, and it has been recognised that transcriptional control working through MYB-related transcription factors is particularly important in plants (Martin and Paz-Ares, 1997; Romero *et al.*, 1998).

What is a MYB-related protein?

The MYB domain is a region of about 52 amino acids that binds DNA in a sequence-specific manner. In c-MYB (historically the prototypic MYB protein) this domain is repeated three times (R1, R2 and R3; Figure 1) and each imperfect repeat adopts a helix-helix-turn-helix conformation to intercalate in the major groove of the target DNA. In plants the predominant family of MYB proteins have two repeats (R2, R3 relative to the repeats of c-MYB; Figure 1). In addition, three-repeat MYBs, closely related to c-MYB, have recently been identified in plants (Ito, in press) together with a growing number of MYB proteins with a single MYB domain (Figure 1).

Diversity in the interaction between MYB proteins and DNA

Structural studies of c-MYB have shown that both R2 and R3 are required for sequence-specific binding, the C-terminal helix of each repeat being the recognition helix for DNA binding. Regularly spaced tryptophan residues (three per repeat) participate in a hydrophobic cluster. It has been suggested that the recognition helix of R3 specifically interacts with the core of the

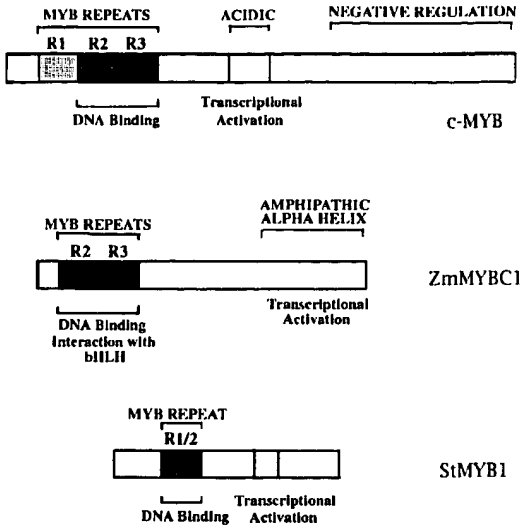

Figure 1. Schematic showing functional domains of prototypic MYB proteins; three MYB repeats (c-MYB), two MYB repeats (C1) and one MYB domain (StMYB1).

recognition sequence, whereas the recognition helix of R2 is involved in less specific interactions with nucleotides peripheral to the core (Ogata *et al.*, 1995). R2R3 repeat MYBs are believed to bind DNA in a similar way.

Because in c-MYB both R2 and R3 are necessary for binding DNA, it is likely that proteins with single MYB domains bind DNA in a different manner. Structural studies on the human telomeric protein, hTRF1, which contains a single MYB domain, suggest that the C-terminal helix is longer than the equivalent helices in the repeats of c-MYB and, consequently, it does not interact with DNA in the same way. In hTRF1 it is proposed that the protein binds DNA in a manner analogous to homeodomain proteins, whose DNA-binding domains also form helix-turn-helix motifs (Nishikawa *et al.*, 1998).

It remains to be seen whether other single MYB-domain proteins bind DNA in a way related to hTRF1, although generally the recognition helices in single MYB domain proteins are not particularly similar to those of c-MYB (relative to their conservation in R2, R3 MYB proteins). It may be that single MYB-domain proteins generally bind DNA in a manner similar to homeodomain proteins and as dimers (either hetero- or homo-dimers), which may have important repercussions for their modes of action and biological functions.

Therefore, members of the superfamily of MYB proteins should be viewed as related principally by their ability to bind DNA, rather than on the basis of their physiological functions. Even given the attribute of DNA binding, there may be major differences in the ways MYB proteins bind DNA. This means that there are different target recognition sites for different groups of MYB proteins, not only between single-domain MYBs and two/three-repeat MYBs, but also within these groupings. Mammalian three-repeat MYBs such as c-MYB, A-MYB and B-MYB and closely related proteins from invertebrates and cellular slime moulds all bind to the cognate site T/CAACG/TGA/C/TA/C/T (MBSI). Some plant two-repeat proteins can recognise this binding site while others cannot. Some of those plant MYB proteins that bind to MBSI will also bind to a second site, TAAC-TAAC (MBSII), which is a sequence recognised by the majority of plant R2R3 MYB proteins (Romero *et al.*, 1998). The group of plant R2R3 MYB proteins that bind preferentially to MBSI (group A) are also more closely related in the primary structure of their DNA-binding domains to the c-MYB family.

Therefore, broad distinctions in target site recognition can be made between MYB proteins on the basis of the structure of their DNA-binding domains, which fall into distinct structural subgroups. However, within the plant R2R3 repeat MYBs, overlaps in binding site recognition have also been reported between members of the different subgroups, and it is likely that most MYB DNA-binding domains have considerable inherent flexibility in their ability to recognise target sites. The binding site preference and affinity of MYB proteins is also likely to be strongly influenced by other protein factors that interact with them. In terms of function, these generalisations mean that MYB proteins belonging to different structural sub-groups are unlikely to have similar functions because of differences in their preferred binding sites. It does not mean the converse, however, i.e. that strong similarity between the DNA-binding domains of MYB proteins implies a commonality of function. Flexibility in recognition, operating through a variety of mechanisms, may mean that proteins very similar in their DNA-binding domains control quite different target genes and therefore have quite distinct physiological functions. For example, within the large family of R2R3 MYB proteins identified in plants three major subdivisions can be made on the basis of the sequence of the DNA-binding domain: subgroup A (whose members are most similar to c-MYB and other animal MYB

proteins), subgroup B which is a relatively small group (4 members in *Arabidopsis*) and subgroup C which encompasses 70 members in *Arabidopsis* (from a total of 83 defined so far), several members of which have been shown to recognise the MBSIIG binding site (T/CACCA/TAC/AC) preferentially (Romero *et al.*, 1998). Members of subgroup C include the *Arabidopsis* gene *AtMYBGL1* which is involved in trichome specification and which may be required for promoting endoreduplication and increasing cell size, the *Antirrhinum* gene *AmMYBMIXTA* which is involved in specifying the formation of conical cells in petal epidermis and which plays no role in cellular outgrowth in *Arabidopsis* and the maize genes *ZmMYBC1*, *ZmMYBPL* and *ZmMYBP*, the *Petunia* gene *PhMYBAN2* and the *Antirrhinum* genes *Am-MYBROSEA* and *AmMYBVENOSA* which are all involved in regulating anthocyanin production. Despite having very similar DNA-binding domains, these subgroup C proteins clearly have distinct physiological functions.

Evolution of MYB DNA-binding domains

The diversity in organisation of MYB domains in proteins is understood best from an evolutionary perspective. An attractive model for evolution of MYB proteins has been presented by Lipsick (1996). This model is based on early (over 1 billion years ago) duplication of the MYB domain to give multiple-repeat MYB proteins followed by later expansion of MYB proteins through duplication of entire genes. This expansion was relatively limited for the three-repeat MYB proteins in animals (vertebrates having just three closely related members, c-MYB, A-MYB and B-MYB). However, in plants the expansion of the R2R3 family was considerable, and there are estimated to be over 100 members of this subfamily in *Arabidopsis*, and equivalent or larger numbers in other plant species.

The Lipsick model for the evolution of MYB repeats remains consistent with new sequence information derived since 1996 (Figure 2). However, it has become apparent that plants have retained genes encoding three-repeat MYBs (Ito, in press) in addition to expanding the R2R3 family (which is believed to have been amplified after loss of R1 from a three-repeat ancestor; Lipsick, 1996).

Interestingly, the function of three-repeat MYBs from plants may also have been conserved since they are thought to be involved in cell cycle control, regulating the expression of cyclins (Ito, in press). In animals, c-MYB, A-MYB and B-MYB are associated with the promotion of cellular proliferation which may operate through control of the cell cycle. The three-repeat MYBs in plants may represent descendants of an ancient member whose structure has been conserved during the evolution of plants and animals, and whose function may be relatively similar in these different kingdoms.

The one MYB domain family has also been expanded in plants; adding to the list including StMYB1, (the first member) are two single-domain MYB proteins from *Arabidopsis* (CCA1 and LHY) believed to operate as oscillators close to or part of the circadian clock governing flowering, leaf movements, photosynthetic gene expression and hypercotyl growth. Two other related single MYB domain proteins have also been identified in *Arabidopsis* but their functions have not been characterised. In addition, the *CAPRICE* gene (*CPC*) encodes a single-domain MYB protein with a role in root hair formation in *Arabidopsis*.

The family of one MYB domain proteins, which recognise G-rich telomeric sequences (TBFs), have been conserved during the evolution of yeasts, animals and plants (Bilaud *et al.*, 1996). This family of MYB proteins may also be able to control transcription; the plant members of the family have been shown to work as transcriptional activators.

Two-repeat MYBs closely related structurally to the product of the *Schizosaccharomyces pombe* gene, *cdc5,* have been conserved in fungi (present in *S. pombe* and *Saccharomyces cerevisiae*), animals and plants (Ohi *et al.*, 1998), and their function is also thought to have been conserved (regulation of G2/M progression in the cell cycle). These cdc5p-related proteins contain a third region just C-terminal to the second MYB repeat which has very weak similarity to classic MYB repeats (it lacks the regularly spaced tryptophan residues, for example) which could represent a highly diverged third repeat.

Another MYB (DMP1), conserved in animals and yeast, binds to D cyclins and is thought to regulate the cell cycle. This is a three-repeat MYB with strongest similarity in its repeats to R1/R2 of c-MYB. The third repeat shows little conservation of structure with the classic c-MYB R3 domain except in maintaining the regularly spaced tryptophan residues. In addition to binding to D cyclins, these MYB proteins can also activate transcription (Hirai and Sherr, 1996).

From studies of amino acid homologies the Lipsick model suggested that the R2R3 MYB-related proteins

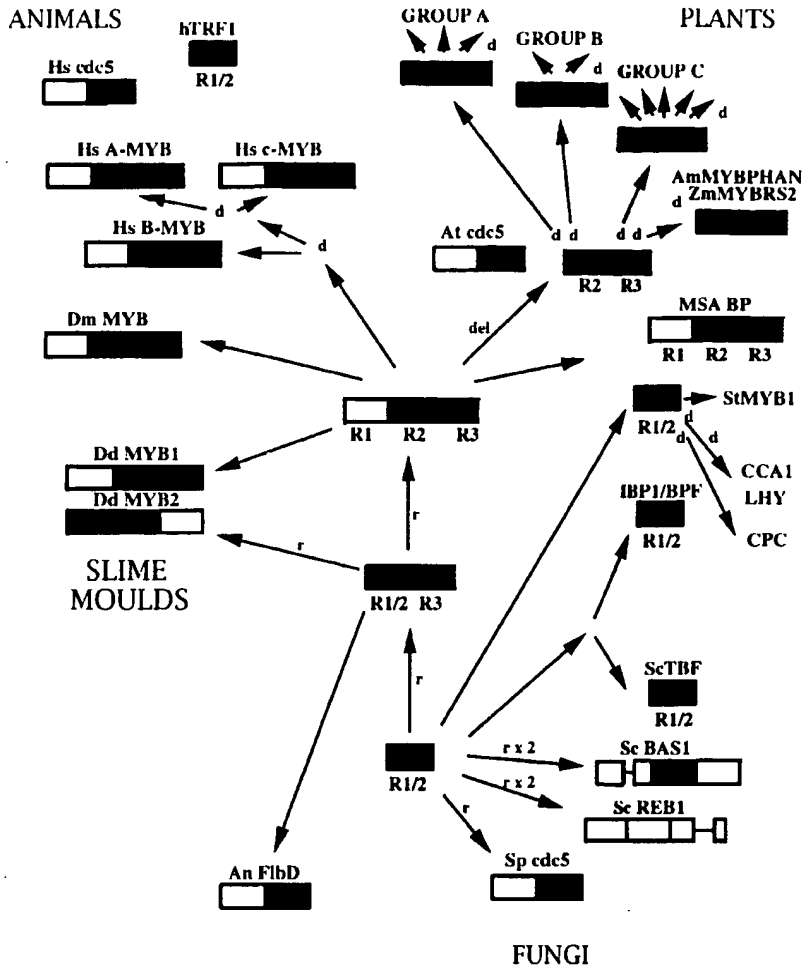

Figure 2. Model for the evolution of MYB proteins in the major eucaryotic kingdoms; slime moulds, animals, plants and fungi (modified from Lipsick, 1996). It is proposed that a single original MYB repeat was replicated (r) to give rise to two- and three-repeat MYB proteins. The majority of plant MYB proteins are thought to have evolved following loss (del) of the first repeat (R1) to give the expanded R2R3 family. Different subgroups within this large family are thought to have arisen by duplication of entire genes (d). Three-repeat MYBs such as MSA BP from tobacco have been maintained in plants (Ito, in press). Other proteins such as the cell cycle regulator Spcdc5 and the telomere-binding proteins have been conserved in animals (Hscdc5 and hTIRF1, respectively), plants (Atcdc5 and IBP1/BPF, respectively) and fungi (Spcdc5 and TBP, respectively), and it seems likely that their functions have been conserved as well. (Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; Dd, *Dictostylium discoides*; An, *Aspergillus nidulans*; Sp, *Schizosaccharomyces pombe*; Sc, *Saccharomyces cerevisiae*; St, *Solanum tuberosum*, Am, *Antirrhinum majus*; Zm, *Zea mays*; At, *Arabidopsis thaliana*).

arose after loss of the sequences encoding R1 in an ancestral three-repeat *MYB* gene. It is likely that, upon loss of R1, several subgroups of genes encoding R2R3 MYB proteins went through selective amplification and subgroup expansion during plant evolution (Lipsick, 1996; Romero *et al.*, 1998; Figure 2). More than 80 *Arabidopsis* genes encoding R2R3 MYBs have been specifically cloned and analysed (Romero *et al.*, 1998). This analysis identified three distinct R2R3 subgroups on the basis of the structure of part of their DNA-binding domain. It seems likely that the three *Arabidopsis* R2R3 MYB subgroups (A, B, C) were derived from different R2R3 ancestors. In addition, the family of R2R3 MYBs characterised by PHANTASTICA (AmMYBPHAN) from *Antirrhinum majus* and ROUGHSHEATH 2 (ZmMYBRS2) from

maize represent an independent subgroup because the sequence of the recognition helix in R3 (which is involved in making base contacts) is very different in AmMYBPHAN and AmMYBRS2 to that found in c-MYB or other R2R3 MYBs of plants (Timmermans *et al.*, 1999). The AmMYBPHAN subgroup of MYBs has been implicated in meristem initiation and control of the dorso-ventral axis in shoot organs (Table 1).

Diversity within the C-terminus of plant MYBs

Most MYB proteins are presumed to be transcriptional activators with activation domains in the region C-terminal to the DNA-binding domain, since c-MYB has an activation domain C-terminal to its DNA-binding domain, which is acidic (Weston, 1998). A few plant MYBs have been tested as transcriptional activators, and most will activate transcription, the activation domain generally being predicted to form an amphipathic α-helix. However sequences in the C-terminal regions of MYB proteins are not strongly conserved, presumably because the structural determinants for activation domains are fairly flexible.

Not all MYB-related proteins need be transcriptional activators. There is circumstantial evidence that some may serve to reduce target gene expression, although no repression domains have yet been defined. In principle, silencing of target gene expression can also be achieved by competition for DNA-binding sites with other transcriptional activators or by binding and titrating out other activators themselves so that activation of transcription is reduced. There are currently no data on the biochemical action of plant MYB proteins with respect to the repression of gene transcription.

A survey of the predicted C-terminal sequences of the *Arabidopsis* R2R3 MYB family revealed 22 different subgroups which showed limited sequence conservation within their C-terminal regions (Kranz *et al.*, 1998). These conserved motifs might represent activation domains, an idea supported by the fact that some are relatively acidic, and that others are rich in amino acids frequently associated with activation domains (glutamine, proline); the sequence requirements for activation domains are known to be flexible. Alternatively, these regions of sequence conservation may represent repression domains or domains for interaction with other transcription factors, although in the only case of plant MYB protein interaction that has been reported, that of the maize protein ZmMYBC1 with bHLH factors B, R, Lc and Sn, the interaction is believed to involve the DNA-binding domain of the MYB protein (Martin and Paz-Ares, 1997).

Multifunctionality of plant MYB proteins

Interest in the diversity of MYB proteins in plants clearly stems from an ultimate interest in their functionality. Faced with the plethora of MYB genes, can conclusions be drawn about their likely functions on the basis of their primary amino acid sequences? In addition, why do plants have so many R2R3 repeat MYBs and how does the individual activity of each member relate to the activity of other members?

Considering all MYB-related proteins, it is clear that there is a wide diversity of function. For example, the primary biological role of the MYB proteins that bind telomeric sequences may be structural, and their role as transcription factors may be secondary. Other single MYB domain proteins clearly function primarily as transcription factors; the biochemical functions of some of these are related to the rhythmic changes in gene expression associated with the circadian clock, while another has a role in root hair formation.

The R2R3 MYB family in plants is large and its functions are diverse (Table 1). The only uniting feature is that most members of the family seem to be involved in 'plant-specific processes' involving control of secondary metabolism or response to secondary metabolites unique to plants or cellular morphogenesis unique to plants (Martin and Paz-Ares, 1997). Although this generalisation may serve to address the question why there was wide expansion of the R2R3 *MYB* gene family in plants it does not help very much in predicting the function of the, as yet, uncharacterised members. Ultimately, our understanding of the range of influence of MYB-related proteins in plants will depend on mutational analysis of each, a task that is currently underway. However, it is still pertinent to ask to what extent homologous functions can be assigned to structurally related MYB proteins from different plant species, and what is the extent of redundancy in *MYB* gene activity.

Are structurally similar MYB proteins functionally homologous?

The best way to consider whether structurally related MYB proteins from different species share homologous functions is to consider the best characterised

Table 1. List of plant MYB-related genes for which function has been assigned. The subgroup for the DNA-binding domain of the R2R3 gene is also listed according to Romero *et al.* (1998).

MYB genes	Biological functions	Species	R2R3 subgroup (Romero *et al.*, 1998)
One-repeat Myb			
StMYB1	Unknown	*Solanum tuberosum*	
LHY	Circadian clock regulation, flowering time	*Arabidopsis thaliana*	
CCA1	Phytochrome & circadian regulation	*Arabidopsis thaliana*	
PcMYB1	Light-dependent activation	*Petroselinum crispum*	
CPC1	Epidermal cell differentiation, root hairs	*Arabidopsis thaliana*	
BPF1	Telomeric DNA binding protein	*Petroselinum crispum*	
IBP1	Telomeric DNA binding protein	*Zea mays*	
R2R3 Myb			
	Phenylpropanoid metabolism		
ZmMYBC1	Anthocyanin	*Zea mays*	Subgroup C
ZmMYBPL	Anthocyanin	*Zea mays*	Subgroup C
ZmMYB1	Anthocyanin	*Zea mays*	Subgroup C
ZmMYB38	Inhibition of C1-mediated activation	*Zea mays*	Subgroup C
PhMYBAN2	Anthocyanin	*Petunia hybrida*	Subgroup C
PhMYB3	Anthocyanin	*Petunia hybrida*	Subgroup C
AmMYB305, 340	Anthocyanin and flavonol	*Antirrhinum majus*	Subgroup C
PsMYB26	Phenylpropanoid regulation	*Pisum sativum*	Subgroup C
ZmMYBP	Phlobaphene	*Zea mays*	Subgroup C
AmMYB308, 330	Phenolic acid	*Antirrhinum majus*	Subgroup C
	Development		
AtMYBGL1	Trichome development	*Arabidopsis thatiana*	Subgroup C
AmMYBMIXTA	Conical cell development	*Antirrhinum majus*	Subgroup C
PhMYB1	Conical cell development	*Petunia hybrida*	Subgroup C
CotMYBA	Trichome development	*Gossypium hirsutum*	Subgroup C
AmMYBPHAN	Dorsoventral determination & growth	*Antirrhinum majus*	AmMYBPHAN subgroup
ZmMYBRS2	PHAN-like, repress *knox* expression	*Zea mays*	AmMYBPHAN subgroup
AtMYB13	Shoot morphogenesis	*Arabidopsis thaliana*	Subgroup C
AtMYB103	Expressed in developing anthers	*Arabidopsis thaliana*	Subgroup C
	Signal transduction		
GAMYB	Gibberellin response	*Hordeum vulgare*	Subgroup B
AtMYB2	Dehydration and ABA regulation	*Arabidopsis thaliana*	Subgroup C
ATR1	Tryptophan biosynthesis	*Arabidopsis thaliana*	Subgroup C
Cpm5, Cpm7, Cpm10	Dehydration and ABA response	*Craterostigma plantagineum*	
	Plant disease resistance		
NtMYB1	TMV, SA-inducible	*Nicotiana tabacum*	
	Cell division		
AtCDC5	Cell cycle regulation	*Arabidopsis thaliana*	
R1RR2R3 Myb			
MSA-binding proteins	Regulation of B-type cyclin genes	*Nicotiana tabacum*	
AtF4D11.7	Unknown	*Arabidopsis thaliana*	
AtF6N23.19	Unknown	*Arabidopsis thaliana*	

subfamily of R2R3 MYBs, those controlling anthocyanin biosynthesis: ZmMYBC1 and ZmMYBPL from maize, PhMYBAN2 from *Petunia* and AmMYBROSEA from *Antirrhinum*. The proteins of these genes are structurally related, especially in their DNA-binding domains which are very similar, but also in their C-terminal sequences. The obvious conclusion is that these genes are structurally and functionally homologous. However it is known that ZmMYBC1 (ZmMYBPL) and PhMYBAN2 do not regulate exactly the same target genes in maize and petunia, in that an2 mutants of petunia are not affected in their expression of some structural genes of anthocyanin biosynthesis, whereas c1 mutants in maize show reduced expression of all the structural genes (Martin and Paz-Ares, 1997). Rosea mutants of *Antirrhinum* show reduced expression of a different subset of structural genes to c1 and an2 mutants, demonstrating that although the effects of loss of gene function are similar in all cases (loss or reduction in pigment production) the biochemical functions of each MYB protein are not precisely homologous. This suggests that those *Arabidopsis* genes encoding similar MYB products (*AtMYB75* and *AtMYB90*) may serve roles in regulating anthocyanin biosynthesis but cannot predict precisely which target genes will be regulated by them.

Do *MYB* genes show extensive redundancy in their functions?

The degree of genetic redundancy within *MYB* gene subfamilies can also be assessed by considering those genes controlling anthocyanin biosynthesis. In maize, the *MYB* genes *ZmMYBC1* and *ZmMYBPL* clearly have the same function: to activate transcription of the structural genes of anthocyanin biosynthesis. The two genes work in different plant tissues: *ZmMYBC1* in the aleurone and some tissues of the flowers, *ZmMYBPL* in the vegetative plant tissues. In this example, paralogous genes, which have most likely arisen by gene duplication, have adopted different expression patterns. It is likely that in other species pigmentation patterns may result from similar activities of paralogous regulatory genes. The maize gene *ZmMYBP* controls phlobaphene biosynthesis in pericarp tissue. Phlobaphenes are derived from the flavonoid pathway that also gives rise to anthocyanins, and ZmMYBP is known to activate some, but not all of the target genes of ZmMYBC1, although it is thought to bind to sites in the promoters of the structural genes with differing affinity to ZmMYBC1. ZmMYBP is quite closely related structurally to ZmMYBC1, suggesting that structurally related proteins (particularly if related in their DNA-binding domains) perform related although not identical functions.

This general idea is supported by studies on the function of some members of R2R3 MYB C-terminal subgroup 4 which share very similar DNA-binding domains and a C-terminal region potentially encoding a zinc-finger motif (Kranz *et al.*, 1998). This C-terminal subgroup is most closely related to the C-terminal subgroups including ZmMYBC1, ZmMYBPL, ZmMYBP and PhMYBAN2 (subgroups 5, 6 and 7; Kranz *et al.*, 1998). It is known that some members of subgroup 4 can regulate expression of genes involved in hydroxycinnamic acid metabolism; another branch of phenylpropanoid metabolism linked to flavonoid metabolism by three common steps at the start of each pathway. Again structural similarity appears to reflect functional similarity, although it is not thought that members of subgroup 4 normally regulate any structural genes in common with ZmMYBC1 or PhMYBAN2.

Even where gene products are closely related structurally and produced in the same species it cannot be assumed that their functions are redundant or that they are paralogous. This is most clearly illustrated by the vertebrate MYB family c-MYB, A-MYB and B-MYB. All three proteins share virtually identical DNA-binding domains, and are known to be able to bind to the same target DNA sequences. However numerous ectopic expression studies and analyses of knock-out mutants of mice suggest that the proteins are not functionally equivalent. C-MYB and A-MYB are structurally most similar, sharing a central activation domain, a region for interaction with the transcription factor CBP, and a C-terminal negative regulatory domain as well as the MYB DNA-binding domain. However c-MYB has another motif in its C-terminal region for interaction with transcriptional co-activators, not found in A-MYB. B-MYB has a C-terminal region involved in cell-cycle-regulated phosphorylation, present in A-MYB but not c-MYB. A-*MYB* is expressed largely in spermatogenic tissue whereas c-*MYB* is expressed in haematopoeietic tissues and epithelium. It is possible that A-MYB and c-MYB, as a result of their structural similarities, are paralogous (in an equivalent way to ZmMYBC1 and ZmMYBPL), regulating cellular proliferation and the commitment to cellular differentiation, through essentially the same target genes, in different tissues.

However, the unique features of each protein argue against identical functions and, rather, suggest overlapping functions, with some but not all target genes in common (Weston, 1998).

B-MYB is less similar to c-MYB than A-MYB. Although B-MYB is expressed generally in tissues it cannot compensate for loss of c-MYB or A-MYB activity (knock-outs in mice give phenotypes) arguing that it is not functionally homologous to either. Lines overexpressing c-MYB and B-MYB also give different phenotypes. B-MYB may regulate the cell cycle through control of cell cycle progression from G1 to S whereas A-MYB may be more closely associated with the control of meiosis. These details suggest that structurally related MYB proteins may share related (overlapping) but non-identical functions (Weston, 1998). It seems likely that such generalisations are also broadly applicable to the MYB subgroups in plants (especially within the R2R3 MYB subgroups).

While the concept of paralogous genes controlling the same functions within a species in different tissues provides an attractive explanation for some of the duplication of R2R3 MYBs in plants, current views on how transcription factors control gene expression suggest that differences in expression pattern may also be instrumental in dictating differences in functionality (Sieweke and Graf, 1998). Transcription factors generally interact with other proteins associated with transcriptional control, and it is unlikely that plant MYBs are exceptions. If they assemble independently of their DNA target sites, into multicomponent complexes, the nature of these complexes will depend upon the interacting proteins available in any particular cell type. Differences in the make-up of these complexes may then result in distinct activities on different target promoters. The idea that differences in protein-protein interactions (dictated in part by differences in gene expression patterns) account for functional differences between very similar transcription factors has already been proposed to account for functional differences between vertebrate MYB proteins. It is equally likely that similar considerations apply to plant MYB transcription factors. There is evidence for this in the case of AmMYBMIXTA, which normally controls the formation of conical cells in petal epidermis. If this gene is ectopically expressed under control of the CaMV 35S promoter in tobacco or *Antirrhinum*, it results in the formation of multicellular trichomes on leaves, a function it does not normally control. This suggests that AmMYBMIXTA can adopt novel functions through changes in its expression pattern, possibly through changing the suite of available interacting proteins. It also suggests that the function of closely related MYB proteins expressed in appropriate cells might be to regulate multicellular trichome formation (Glover *et al.*, 1998).

Conclusions

MYB genes have expanded and diversified their functions during the evolution of flowering plants, and now regulate many different aspects of metabolism and development. The range of involvement of MYB transcription factors in controlling different aspects of plant gene expression is by no means fully characterised. While it is clear that a significant number of these MYB transcription factors are involved in the detailed regulation of secondary metabolism (particularly phenylpropanoid metabolism), it is unlikely that many of the MYB proteins operating within a single species are truly redundant in their functions. In addition, the difficulty in establishing functional homology between structurally homologous proteins from different plant species suggests that variations in the activity of these regulatory proteins may make a major contribution to the variation in traits between species.

Acknowledgements

We would like to thank all members of the 'MYB Consortium', particularly Javier Paz-Ares, Chiara Tonelli, Mike Bevan, Bernd Weisshaar and Sjef Smeekens, for stimulating and enlightening discussions on the subject of MYB regulatory proteins in plants. The 'MYB Consortium' was funded by the European Union, Framework IV, Contract CT95-0129.

References

Bilaud, T., Koering, C.E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S.M. and Gilson, E. 1996. The telobox, a MYB-related telomeric DNA binding motif found in proteins from yeast, plants and human. Nucl. Acids Res. 24: 1294–1303.

Glover, B.J., Perez-Rodrigues, M. and Martin, C. 1996. Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor. Development 125: 3497–3508.

Hirai, H. and Sherr, C.J. 1996. Interaction of D-type cyclins with a novel MYB-like transcription factor, DMP1. Mol. Cell. Biol. 16: 6457–6467.

Kranz, H.D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R., Petroni, K., Urzainqui, A., Bevan, M., Martin, C., Smeekens, S., Tonelli, C., Paz-Ares, J. and Weisshaar, B. 1998. Towards functional characterisation of members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. Plant J. 16: 263–276.

Lipsick, J.S. 1996. One billion years of MYB. Oncogene 13: 223–235.

Martin, C. and Paz-Ares, M.Y.B. 1997. transcription factors in plants. Trends Genet. 13: 67–73.

Nishikawa, T., Nagadoi, A., Yoshimura, S., Aimoto, S., Nishamura, Y.: Solution structure of the DNA-binding domain of human telomeric protein hTRF1. Structure 6: 1057–1065.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimolo, S., Ametani, Y., Hirata, Z., Sarai, A., Ishii, S. and Nishimura, Y. 1995. Comparison of the free and DNA-complexed forms of the DNA-binding domain of c-MYB. Nature Struct. Biol. 2: 309–320.

Ohi, R., Feoktistova, A., McCann, S., Valentine, V., Look, A.T., Lipsick, J.K. and Gould, K.L. 1998. MYB-related *Schizosaccharomyces pombe* cdc5p is structurally and functionally conserved in eucaryotes. Mol. Cell. Biol. 18: 4097–4108.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. and Saedler, H. 1987. The regulatory *c1* locus of *Zea mays* encodes a protein with homology to *myb* oncogene products and with structural similarities to transcriptional activators. EMBO J. 6: 3553–3558.

Sieweke, M.H. and Graf, T. 1998. A transcription factor party during blood cell differentiation. Curr. Op. Gene Devel. 8: 545–541.

Romero, I., Fuertes, A., Benito, M.J., Malpica, J.M., Leyva, A. and Paz-Ares, J. 1998. More than 80 *R2R3-MYB* regulatory genes in the genome of *Arabidopsis thaliana*. Plant J. 14: 273–284.

Timmermans, M.C.P., Hudson, A., Becraft, P.W. and Nelson, T. 1999. ROUGHSHEATH2: A MYB protein that represses *knox* homeobox genes in maize lateral organ primordia. Science 284: 151–153.

Weston, K. 1999. MYB proteins in life, death and differentiation. Curr. Op. Gene Devel. 8: 76–81.

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.